DOCUMENT RESUME

ED 263 119 TM 850 504

AUTHOR Eignor, Daniel R.
TITLE An Investigation of the Feasibility and Practical Outcomes of Pre-Equating the SAT Verbal and Mathematical Sections.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-85-10
PUB DATE Mar 85
NOTE 124p.; Part of this paper was presented at the Annual Conference of the American Educational Research Association (Montreal, Canada April 11-14, 1983).
PUB TYPE Reports - Research/Technical (143) -- Statistical Data (110)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Equated Scores; Estimation (Mathematics); Feasibility Studies; Higher Education; Item Analysis; *Latent Trait Theory; Mathematical Models; Mathematics Tests; Pretesting; Raw Scores; *Scores; Secondary Education; Test Format; Test Items; True Scores; Verbal Tests
IDENTIFIERS Pre Equating (Tests); *Scholastic Aptitude Test; Test Security

ABSTRACT

The feasibility of pre-equating, or establishing conversions from raw to scaled scores through the use of pretest data before operationally administering a test, was investigated for the Scholastic Aptitude Test (SAT). Item-response theory based equating methods were used to estimate item parameters on SAT pretest data, instead of using final form SAT test results which were derived from multiple pretests. The verbal and mathematical sections of two SAT forms were calibrated from pretest data, pre-equated to existing SAT forms, and the pre-equating results were compared to final form item response theory equating and conventional linear equating methods. The results indicated that pre-equating of the two verbal test forms varied considerably. The equating for one form was found to be reasonably acceptable, while for the other it was marginally acceptable or unacceptable. This was explained by the location of reading passages and reading comprehension items at the end of the tests. The equating results were fairly similar for both forms of the mathematical test. They were considered only marginally acceptable, or perhaps unacceptable. Contributing reasons for this were not clear. Item difficulty was not consistent between pretests and the final test form. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-85-10

ED263119

RESEARCH REPORT

AN INVESTIGATION OF THE FEASIBILITY AND PRACTICAL OUTCOMES OF PRE-EQUATING THE SAT VERBAL AND MATHEMATICAL SECTIONS

Daniel R. Eignor

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Widenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Educational Testing Service
Princeton, New Jersey
March 1985

TM 850 504

An Investigation of the Feasibility and Practical Outcomes of Pre-equating the SAT Verbal and Mathematical Sections[1,2,3]

Daniel R. Eignor
Educational Testing Service

[1]This study was supported by The College Board through Joint Staff Research and Development Committee funding.

[2]The SAT-verbal portion of this study was presented at the annual meeting of AERA, Montreal, 1983.

[3]The author would like to acknowledge the technical assistance of Martha Stocking and Nancy Wright and the programming assistance of Inge Stiebritz and Karen Carroll in performing this study.

Copyright © 1985, Educational Testing Service. All rights reserved.

An Investigation of the Feasibility and Practical Outcomes of Pre-equating the SAT Verbal and Mathematical Sections

Abstract

Test disclosure legislation, enacted and presently being considered in several states, has serious implications for testing programs relying on conventional equating methods to put new forms of exams on scale. Certain of the problems presented by this legislation can be circumvented by applying item response theory (IRT) based equating methods. Many of the problems can be circumvented if the IRT equating performed is pre-equating, that is, establishing conversions from raw to scaled scores, through the use of pretest data, before the time the new test is administered operationally. The purpose of this study was to determine the extent to which item parameters estimated on SAT-verbal and SAT-mathematical pretest data can be used for equating purposes in a situation where intact final form SAT testing data has normally been used. More specifically, the items that appear in any final SAT form come from multiple pretests, and to the extent that the item statistics are sensitive to the context or position in which the items appear, there may be a lack of fit between the equating based on pretest data and final form data. In this study, the verbal and mathematical sections of two SAT forms were calibrated from pretest data, pre-equated to existing SAT forms, and then the results of the pre-equating compared in a number of ways to final form IRT equating and conventional linear equating methods.

An Investigation of the Feasibility and Practical Outcomes of Pre-equating the SAT Verbal and Mathematical Sections

Daniel R. Eignor
Educational Testing Service

Introduction

The current thrust of research devoted to the applications of item-response theory (IRT) has generated an active interest in the use of IRT methods in the solution of score equating problems (see Cook and Eignor, 1983). Because of the special properties of test data characterized by IRT models, users are often able to solve problems not amenable to traditional equating methods. For other situations, IRT equating offers an alternative against which to evaluate traditional methods. In addition, a number of other important outcomes accrue from the use of IRT for equating tests; among these are 1) improved equating, including better equating at the ends of the scale where important decisions are often made, 2) greater test security through less dependence on items in common with a single old form, 3) easier re-equating should items be deleted, and 4) the possible reduction of bias or drift in equating introduced when traditional methods are used over time in certain situations, most notably when the equating samples for the old and new forms are not random samples from the same population.

While the above listed outcomes accrue as the result of the application of any IRT equating method, if the test forms to be equated can be pre-equated using IRT methods, a number of additional advantages result. Pre-equating refers to the process of establishing conversions from raw to scaled scores prior to the time the new test is administered operationally. The process depends on the adequate pretesting of a pool of items from which

the new test will be built, the calibration of these items using IRT methods, and the utilization of a linking scheme to place the IRT parameters from the pretested items on the same scale. Among the additional advantages offered by IRT pre-equating are the following: 1) Since equating using IRT pre-equating methods is possible prior to the actual administration of the test, new forms can be introduced at low volume special administrations, a particular problem if traditional methods are used; 2) since pre-equating permits linkages to many old forms, it is the most likely of any equating method to yield acceptable results should testing legislation mandate the disclosure of pretest or equating items; 3) pre-equating would allow more time to do reasonableness and quality control checks, which are normally done in a hurried fashion due to score reporting deadlines; and 4) pre-equating would actually permit a reduction in the usual score reporting cycle while simultaneously allowing more time to do the equating itself. In short, the listed advantages that can potentially result from the use of IRT pre-equating build a strong case for investigation of the feasibility of application of this method. In this report, the applicability of IRT pre-equating to the Scholastic Aptitude Test (SAT) verbal and mathematical tests is considered.

Problem and Purpose

To date, investigations of the feasibility of pre-equating using IRT for tests developed and administered by Educational Testing Service for The College Board have been done using data from the Test of Standard Written English (TSWE) (Bejar and Wingersky, 1982). The Bejar and Wingersky study

indicated some discrepancies between pre-equating results and the results from traditional equating in situations where traditional equating was a reasonable procedure. The calibration system used for pre-equating TSWE was considerably different, however, from any system that could be devised for pre-equating the SAT. Thus, although the results of the TSWE pre-equating study were not altogether promising, there is little reason to suggest that these results are generalizable to pre-equating the SAT. For this reason, it was deemed important to investigate the feasibility of pre-equating the SAT using an appropriate calibration system, such as that devised for this study.

The purpose of this study was to determine the extent to which item parameters estimated on SAT-verbal and SAT-mathematical pretest data can be used for equating purposes in a situation where intact final form SAT testing data have normally been used. The items that appear in any final SAT form come from multiple pretests and to the extent that the item parameter estimates are sensitive to the context or position in which the items appear, there may be differences between these parameter estimates and parameter estimates generated using data from the actual final form administration, resulting in a discrepancy between equating based on pretest item parameter estimates and intact final form item parameter estimates. More specifically, in the study, verbal and mathematical items appearing in two final SAT forms, 3ASA3 and 3BSA3, were calibrated almost completely from pretest data. (See the section "IRT Calibration Design and Linkage System".) Elaborate linkage systems, quite representative of the systems that would exist were pre-equating to be considered for operational use,

were devised for the verbal and mathematical items and utilized to get parameter estimates for these items, contained in multiple pretests, on the same scale. The two verbal sections, one from 3ASA3 and the other from 3BSA3, were both part of one linkage system and the two comparable mathematical sections were part of the other.

The effects of using the parameter estimates, obtained from the pretest data, on the equating process were evaluated in the following way. Each of the SAT-verbal and SAT-mathematical final forms under study, when administered for the first time operationally, had been equated by conventional linear methods to two different old forms and the results of the equatings averaged. These equatings were redone using item parameter estimates based on the pretest data and item parameter estimates generated from the intact final form administration. In each case, IRT true-score equating was performed. For each form, the IRT equating based on pretest statistics was compared to the IRT equating based on intact final form data and the linear equating used operationally when each form was put on scale. IRT equating based on intact final form data and linear equating results were used as criteria in this study for the following reasons: (1) In recent IRT equating feasibility studies (Petersen, Cook, and Stocking, 1983; Kingston and Dorans, 1982), it has been demonstrated that intact form IRT true-score equating is a viable equating method for aptitude test data; and (2) the linear methods actually performed to put the forms on scale operationally have undergone many years of scrutiny through their use for operational score reporting purposes.

This study used two SAT-verbal forms and two SAT-mathematical forms so that the consistency of results could be assessed. This should provide the basis for drawing stronger conclusions about the feasibility of pre-equating the SAT than had the replication not taken place.

Methodology

Description of Tests

Test booklets containing SAT forms such as those used in this study consist of six 30-minute sections: two SAT-verbal sections, two SAT-mathematical sections, one Test of Standard Written English (TSWE), and one variable section. The two SAT-verbal sections together comprise the overall SAT-verbal test or form and the two SAT-mathematical sections together comprise the overall SAT-mathematical test or form. All examinees at a given administration take the same test sections except for the variable section, where different subsamples of the total group receive different variable sections. The variable section consists of either one of two verbal or mathematical common item equating sections (anchor tests) or one of a number of verbal, mathematical, or TSWE pretests. In this study, data from all sections except the Test of Standard Written English and variable section TSWE pretests were used. The samples used for calibration purposes in the verbal portion of the study either took the two verbal sections and one of the verbal common item equating sections or the two verbal sections and one of the verbal pretests. The samples used for calibration purposes in the mathematical portion of the study took either the two mathematical sections and one of the mathematical common item equating sections or the two mathematical sections and one of the mathematical pretests.

The two SAT-verbal sections together contain a total of 85 five-choice items (45 items in one section, 40 items in the other section) comprised of 25 antonyms, 20 analogies, 15 sentence completions, and 5 reading passages each of which is followed by 5 items based on the passage. The verbal common item equating sections contain 40 items (10 of each type); these sections are built to be as parallel as possible to the 40 item SAT-verbal section. The verbal pretest sections consist of either 45 or 40 items and are built to be as parallel as possible to the comparable length SAT-verbal sections.

The two SAT-mathematical sections together contain a total of 60 four- and five-choice items (35 items in one section, 25 items in the other section) comprised of 40 five-choice regular mathematics items and 20 four-choice quantitative comparison items. The mathematical common item equating sections each contain 25 regular mathematics items and are built to be as parallel as possible to the 25 item SAT-mathematical section, which also contains regular mathematics items. The mathematical pretest sections contain either 35 or 25 items and are built to be as parallel as possible to the comparable length SAT-mathematical sections.

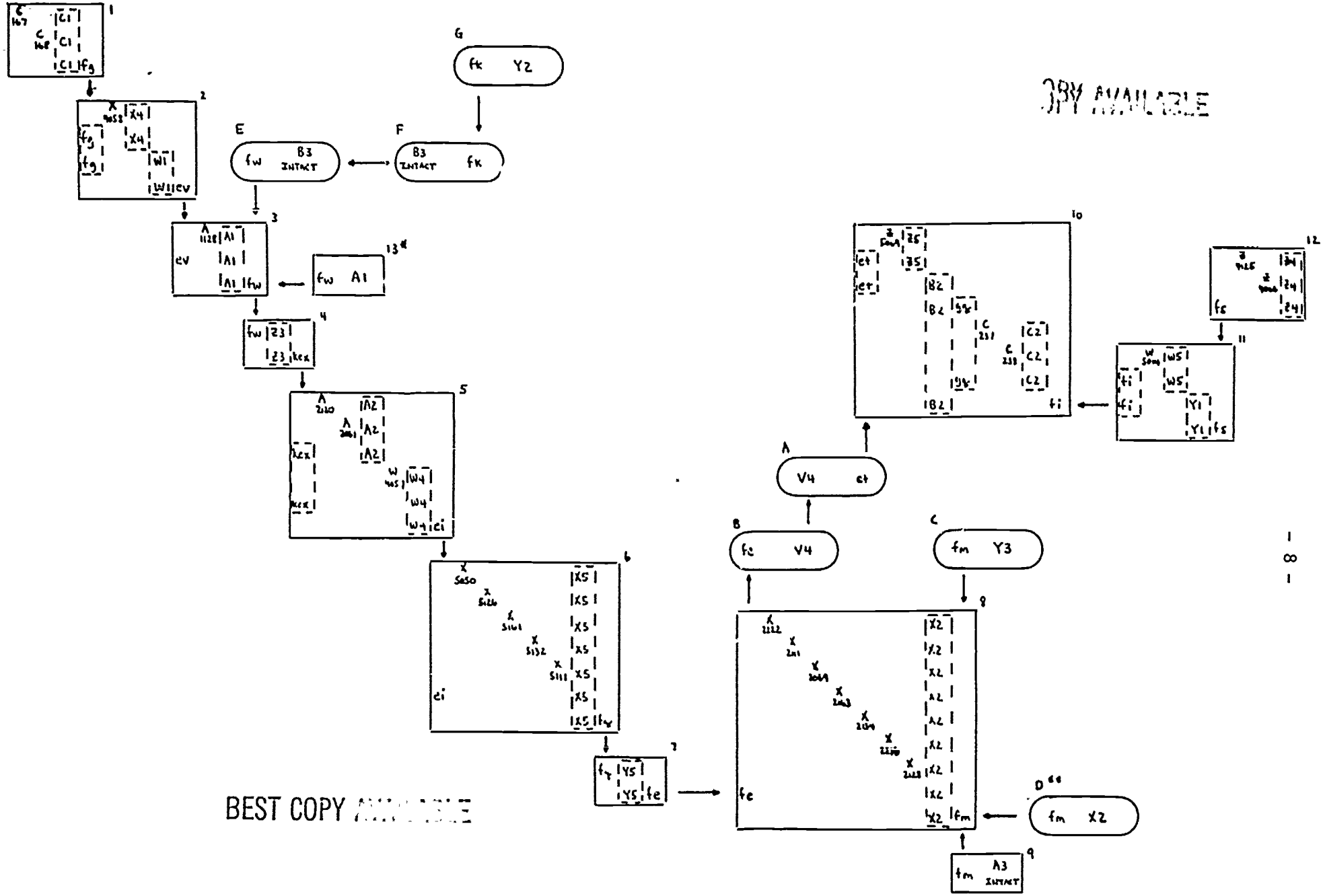
Prior to 1982, raw scores on SAT-verbal (the overall 85 items) and SAT-mathematical (the overall 60 items) were typically transformed to scaled scores on the College Board 200 to 800 scale via linear equating methods. (Separate 200 to 800 scales exist for SAT-verbal and SAT-mathematical.) Since January of 1982, IRT true-score equating using intact final form data has been used to put raw scores on scale. SAT-verbal and mathematical raw scores are obtained scores that have been corrected for guessing. Raw scores are computed by the formula $R-\frac{W}{k}$ where R is the number of correct responses, W is the number of incorrect responses, and (k+1) equals the number of choices per item.

Item Calibration Design and Linkage System

Pretest items corresponding to the two verbal and mathematical sections of two forms of the SAT, 3ASA3 and 3BSA3, were calibrated and placed on a common scale through elaborate linkage systems, one for verbal and one for mathematical, which utilized data on overlapping items from the administration of other intact final forms with either pretest sections or common item equating sections. The calibration linkage system, involving the pretests, final forms, and equating sections for SAT-verbal is depicted in Figure 1; the comparable linkage system for SAT-mathematical is depicted in Figure 2. Responses from randomly selected samples of approximately 3000 examinees taking each pretest-final form combination and approximately 2700 taking each final form-equating section combination were used for calibration purposes.

Each box in Figure 1 or Figure 2 represents a separate calibration (computer run). The dotted-line boxes within the larger boxes indicate the overlapping items that were used to place parameter estimates on the same scale within a single calibration run. The directional arrows between the boxes indicate that a scaling program (described in a later section of this paper) was run to place parameter estimates from the separate calibration runs on the same scale. For SAT- verbal, it should be noted that all items contained in each 40 item equating section appearing in Figure 1 were calibrated; however this was not the case for all items in each pretest of final form. In order to reduce calibration costs, only the 40 item section of SAT-verbal forms used for linking purposes and only the 170[1] (85 items X 2 forms) verbal pretest items which eventually appeared in final forms 3ASA3

[1]Verbal pretest data did not exist for 8 of the 85 items in Form 3ASA3. Therefore, final form data had to be used in the calibration system. This data was obtained from calibration run number 9 in Figure 1.



BEST COPY AVAILABLE

Figure 1. Verbal pre-equating calibration and linking plan. Upper-case letters followed by one digit designate intact SAT final forms. (In the diagram, all intact SAT final form designations are abbreviated, A3 is an abbreviation for 3ASA3, etc.) Upper-case letters followed by three or four digits designate pretest forms. Lower-case letters designate common item equating sections. Boxes (numbered 1-13) indicate separate LOGIST calibration runs. Dotted boxes within the larger boxes indicate overlapping items that were used to place parameter estimates on the same scale within a LOGIST run. Ovals (lettered A-G) indicate forms and equating sections for which parameter estimates already exist from SAT IRT Drift Study. Arrows indicate direction in which scaling (linking) runs took place.

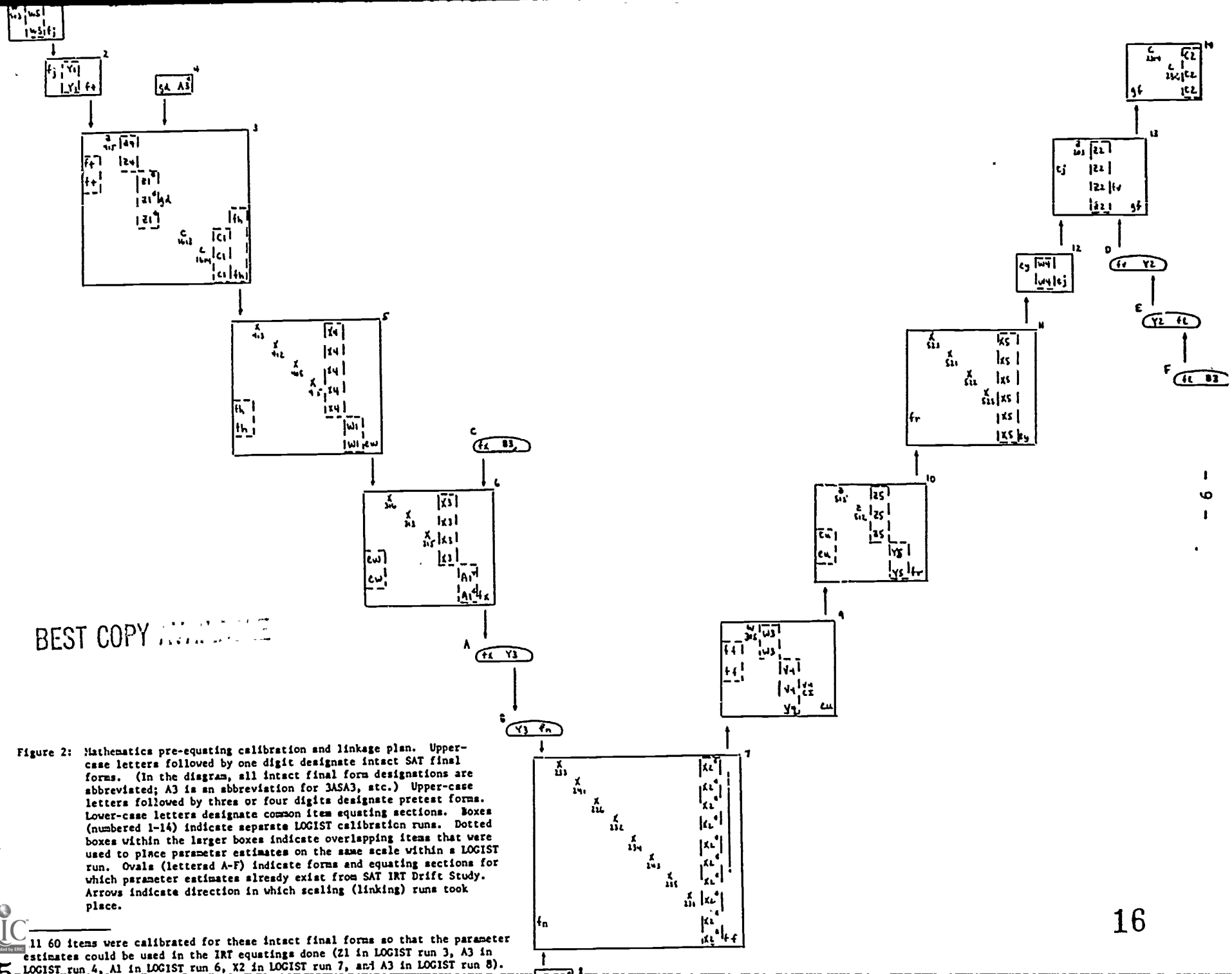


Figure 2: Mathematics pre-equating calibration and linkage plan. Upper-case letters followed by one digit designate intact SAT final forms. (In the diagram, all intact final form designations are abbreviated; A3 is an abbreviation for 3ASA3, etc.) Upper-case letters followed by three or four digits designate pretest forms. Lower-case letters designate common item equating sections. Boxes (numbered 1-14) indicate separate LOGIST calibration runs. Dotted boxes within the larger boxes indicate overlapping items that were used to place parameter estimates on the same scale within a LOGIST run. Ovals (lettered A-F) indicate forms and equating sections for which parameter estimates already exist from SAT IRT Drift Study. Arrows indicate direction in which scaling (linking) runs took place.

---

11 60 items were calibrated for these intact final forms so that the parameter estimates could be used in the IRT equatings done (Z1 in LOGIST run 3, A3 in LOGIST run 4, A1 in LOGIST run 6, X2 in LOGIST run 7, and A3 in LOGIST run 8).

and 3BSA3 were calibrated. Table 1 contains the total number of verbal items and also the total number of examinees responding to the items for each of the 13 SAT-verbal calibration runs. Table 2 lists the number of verbal pretest items calibrated in each of the runs. For SAT-mathematical, all items contained in each 25 item equating section appearing in Figure 2 were calibrated; however this was not the case for all items in each pretest or final form. In order to reduce calibration costs, only the 35 item sections of SAT-mathematical forms used for linking purposes and only the 120[1] (60 items x 2 forms) mathematical pretest items which eventually appeared in final forms 3ASA3 and 3BSA3 were calibrated. Table 3 contains the total number of mathematical items and also the total number of examinees responding to each of the 14 SAT-mathematical calibration runs. Table 4 lists the number of mathematical pretest items calibrated in each of the runs.

Further reduction in the costs of this study were made possible by using existing parameter estimates from the SAT IRT Scale Drift Study (Petersen, Cook, and Stocking, 1983) whenever possible. Also, certain final form-equating section combinations from the Scale Drift Study (labeled C-G in Figure 1 and C-F in Figure 2) and certain final form-equating section calibration runs (numbered 9 and 13 in Figure 1 and numbered 4 and 8 in Figure 2) were linked into the overall calibration linking system, though they were not essential to getting the pretest parameter estimates on the same scale. This was done for equating purposes, and will be described in a later section.

---

[1]Mathematical pretest data did not exist for two of the 60 items in Form 3ASA3. Therefore, final form data had to be used for calibration purposes for one of these items and data on the other item as it appeared in an equating section had to be used.

Table 1

Total Number of Items and Total Number of Examinees
for each of the SAT-verbal LOGIST Calibration Runs

LOGIST Calibration[1] Run Number	Total Number of Items Calibrated	Number of Pretest Items Calibrated	Number of Equating Section Items Calibrated	Number of SAT-verbal Section Items Calibrated	Total Numbers of Examinees
1	135	55	40	40	8,459
2	162	2	80	80	8,519
3	121	1	80	40	7,964
4	120	-	80	40	6,181
5	174	14	80	80	14,069
6	132	12	80	40	22,922
7	120	-	80	40	5,123
8	137	17	80	40	25,778
9	125	-	40	85	2,777
10	298	58	120	120	20,460
11	161	1	80	80	10,347
12	82	2	40	40	8,146
13	125	-	40	85	2,754
	1,892	162[2]	920	810	143,499

[1]LOGIST run number refers to identification scheme in Figure 1.

[2]Pretest data did not exist for 8 of the 85 items in 3ASA3, and hence, final form data had to be used for calibration purposes. Thus only 162 of the total 170 pretest items were calibrated.

BEST COPY AVAILABLE

Table 2

Number of Items Calibrated from each SAT-verbal Pretest Form

Pretest Form	LOGIST[1] Run No.	Total No. of Items Calibrated	No. of Items in 3ASA3	No. of Items in 3BSA3
C167	1	27	13	14
C168	1	28	16	12
X4058	2	2	-	2
A1128	3	1	-	1
A2120	5	7	-	7
A2061	5	4	-	4
W4057	5	3	3	-
X5050	6	3	-	3
X5126	6	2	-	2
X5161	6	1	-	1
X5132	6	1	-	1
X5111	6	5	-	5
X2222	8	2	1	1
X2111	8	1	-	1
X2069	8	1	-	1
X2163	8	2	1	1
X2134	8	4	4	-
X2216	8	1	1	-
X2128	8	6	6	-
Z5069	10	1	-	1
C237	10	29	15	14
C238	10	28	14	14
W5014	11	1	1	-
Z4125	12	1	1	-
Z4066	12	1	1	-
Totals		162[2]	77[2]	85

[1]LOGIST run number refers to the identification scheme in Figure 1.

[2]Pretest data did not exist for 8 of the 85 items in 3ASA3, and hence, final form data had to be used for calibration purposes. Thus, only 77 (of 85) pretest items were calibrated for 3ASA3 and 162 (of 170) for both forms.

Table 3

Total Number of Items and Total Number of Examinees
for each of the SAT-math LOGIST Calibration Runs

LOGIST Calibration[1] Run Number	Total Number of Items Calibrated	Number of Pretest Items Calibrated	Number of Equating Section Items Calibrated	Number of SAT-math Section Items Calibrated	Total Numbers of Examinees
1	61	1	25	35	5,441
2	85	-	50	35	4,692
3	239	35	75	129	22,071
4	85	-	25	60	2,773
5	125	4	50	70	19,007
6	151	6	50	95	16,195
7	128	19	49	60	25,291
8	84	-	24	60	2,744
9	121	1	50	69	13,735
10	127	7	50	70	13,281
11	92	7	50	35	16,594
12	85	-	50	35	5,432
13	110	1	75	35	7,838
14	97	37	25	35	7,981
	1,590	118[2]	648	823	163,075

[1]LOGIST run number refers to identification scheme in Figure 2.

[2]Pretest data did not exist for two of the 60 items in 3ASA3. Final form data had to be used for calibration purposes for one of these items and data on the other item as it appeared in an equating section had to be used.

Table 4

Number of Items Calibrated from each SAT-math Pretest Form

Pretest Form	LOGIST[1] Run No.	Total No. of Items Calibrated	No. of Items in 3ASA3	No. of Items in 3BSA3	Pretest Form	LOGIST[1] Run No.	Total No. of Items Calibrated	No. of Items in 3ASA3	No. of Items in 3BSA3
W503	1	1	-	1	X234	7	3	3	-
Z415	3	1	-	1	X243	7	4	4	-
C1613	3	18	10	8	X235	7	1	-	1
C1614	3	16	7	9	X231	7	1	-	1
X413	5	1	-	1	W305	9	1	-	1
X412	5	2	-	2	Z515	10	3	1	2
X415	5	1	1	-	Z512	10	4	3	1
X316	6	2	2	-	X523	11	3	-	3
X313	6	2	-	2	X521	11	2	-	2
X315	6	2	-	2	X522	11	1	-	1
X233	7	4	3	1	X525	11	1	-	1
X241	7	2	2	-	Z203	13	1	-	1
X226	7	1	1	-	C2314	14	21	10	11
X232	7	3	2	1	C2318	14	16	9	7
					Totals		118[2]	58[2]	60

[1]LOGIST run number refers to the identification scheme in Figure 2.

[2]Pretest data did not exist for two of the 60 items in 3ASA3. Final form data had to be used for calibration purposes for one of these items and data on the other item as it appeared in an equating section had to be used. Thus, only 58 (of 60) pretest items were calibrated for 3ASA3 and 118 (of 120) for both

IRT Model and Item Calibration

Item response theory (IRT) assumes that there is a mathematical function which relates the probability of a correct response or an item to an examinee's ability. (See Lord, 1980, for a detailed discussion.) Many different mathematical models of this functional relationship are possible. The model chosen for this study was the three-parameter logistic model. In this model, the probability of a correct response to item i, $P_i(\theta)$, is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-1.702\, a_i(\theta - b_i)}}, \qquad (1)$$

where a_i, b_i, and c_i are three parameters describing the item and θ represents an examinee's ability. These parameters have specific interpretations: b_i is the point on the θ metric at the inflection point of $P_i(\theta)$ and is interpreted as the item difficulty; a_i is proportional to the slope of $P_i(\theta)$ at the point of inflection and represents the item discrimination; and c_i is the lower asymptote of $P_i(\theta)$ and represents a pseudo-guessing parameter.

The item parameters and examinee abilities for this study were calibrated using the program LOGIST (Wingersky, Barton, and Lord, 1982; Wingersky, 1983). The estimates are obtained by a (modified) maximum likelihood procedure with special procedures for the treatment of omitted items (see Lord, 1974).

LOGIST requires as input the responses to a set of items from a group of examinees, coded to reflect items answered correctly, incorrectly, omitted, and not reached. In addition, the user may specify certain restrictions on the data and parameters in order to speed convergence of the iterative

BEST COPY AVAILABLE

procedure. The major restrictions specified for the study for most of the LOGIST computer runs were:

1. examinees who answered less than one-third of the items were not used,
2. a's were restricted to a range of .01 to 1.75,
3. c's were restricted to a range of .0 to the lesser of .50 or .75(p+), and
4. θ's were restricted to a range of -7.0 to 5.0.

LOGIST produces as output estimates of the a, b, and c for each item, and θ for each examinee.

Thirteen separate LOGIST runs were necessary to calibrate the verbal pretest items, verbal final form and equating section items used for linking purposes, and the verbal final forms to be used for equating purposes. These LOGIST runs are numbered 1-13 in Figure 1. Fourteen separate LOGIST runs were necessary to calibrate the mathematical pretest items, mathematical final form and equating section items used for linking purposes, and the mathematical final forms used for equating purposes. These LOGIST runs are numbered 1-14 in Figure 2. Each of the separate LOGIST runs represented in Figure 1 or Figure 2 generated item parameter estimates on the particular scale defined by the ability distribution of the group of examinees used in the calibration, and hence, a scaling program had to be run to put parameter estimates from the separate LOGIST runs on a common scale. This scaling program also had to be run to put the final form-equating section combinations from the SAT IRT Scale Drift Study (Petersen, et al, 1983) on the common scale. LOGIST run 10 in Figure 1 was chosen as the base form for scaling purposes for SAT-verbal because it contained an SAT-verbal form and equating section which are in common with a

partial pre-calibration linkage system recently devised (Cook and Petersen, 1982) for possible future operational SAT use. LOGIST run 14 in Figure 2 was chosen as the base form for scaling purposes for SAT-mathematical for the same reason.

Item Parameter Scalings

The scalings just referred to are indicated by the directional arrows in Figures 1 and 2 (and also Figures 3 and 4, to be discussed in the following section). A recently devised scaling method (Stocking and Lord, 1983) was used in the study. Briefly, the method works as follows. Letting b, a, and c denote item difficulty, discrimination, and lower asymptote parameters, a linear transformation of the form

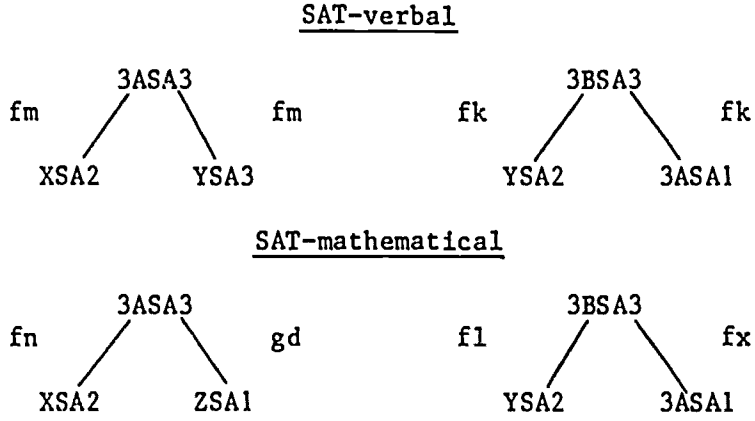
$$b_T = rb + m,$$
$$a_T = a/r \qquad (T = \text{transformed}) \qquad (2)$$

is found which places new form item parameters on the base form scale. The r and m of this transformation are chosen to minimize the average squared difference between true scores on the common item set for a particular group of examinees who have taken the base form. It should be noted that $c_T = c$, so that there is no necessity to transform lower asymptote parameters. This method implicitly makes use of data from all the parameters characterizing an item because true scores are used in the minimization process.

Equating Design

Operationally, the overall verbal section of 3ASA3 and the overall verbal section of 3BSA3 were each linearly equated to two old SAT forms and the results averaged. The overall mathematical section of 3ASA3 and the overall mathematical section of 3BSA3 were also each linearly equated to two

old SAT forms and the results averaged. These equatings can be used as a means for evaluating the effects of using items calibrated from pretest data in the equating process. The following diagram depicts the actual equatings that took place, and the common item sections used for the equatings.



For each of the eight equatings depicted, four for SAT-verbal and four for SAT-mathematical, IRT true-score equating, to be described in detail in the next section, was done three different ways. The first way, referred to as IRT pre-equating, involved the use of item parameter estimates based on pretest items which constitute 3ASA3 and 3BSA3, while the other two ways (both used as criteria to evaluate the IRT pre-equating) involved the use of item parameter estimates based on data collected when 3ASA3 and 3BSA3 were administered as final forms in an intact fashion. The second and third ways differ in the following fashion. In one situation, referred to as intact form calibration system equating, item parameter estimates for 3ASA3, 3BSA3, and the old forms to which they were equated were placed on the same scale, which is essential for IRT equating, by being linked into the overall calibration and linking plans shown in Figure 1 and Figure 2. In this situation, the forms to be equated were linked indirectly through multiple scaling runs applied to a number of intervening LOGIST runs which contain

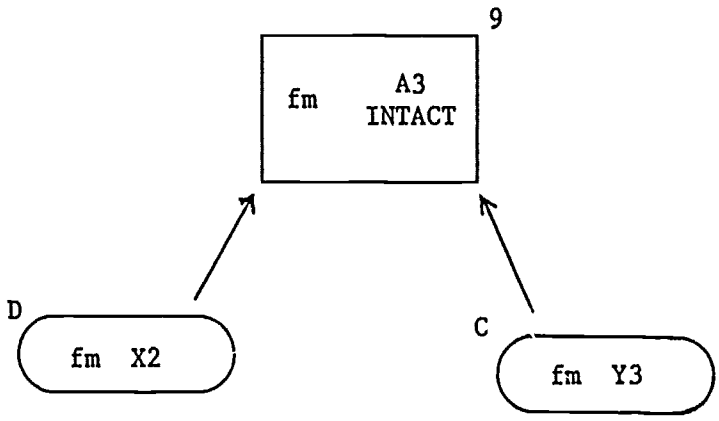
multiple final forms and equating sections. This was done in an attempt to simulate conditions of one possible model under which intact final form IRT equating might take place for the SAT in the future. In the other case, referred to as intact form direct link equating, parameter estimates for the new (3ASA3 and 3BSA3) and old forms to be equated were linked directly through common equating sections, using the scaling procedure described in the previous sections. These linkings are depicted in Figure 3 for SAT-verbal and Figure 4 for SAT-mathematical.

Equating Methods

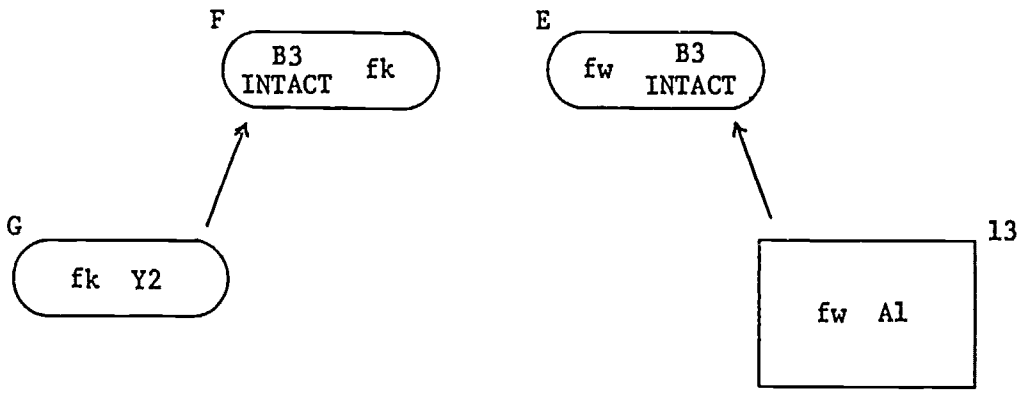
Linear equating methods produce an equating transformation of the form $T(x) = Ax + B$, where T is the equating transformation, x is the test score to which it is applied, and A and B are parameters estimated from the data. The Tucker, Levine Equally Reliable, and Levine Unequally Reliable linear equating models (Angoff, 1971, pp.579-583) are the models that have been used until 1982 for equating SAT-verbal and SAT-mathematical. Choice of which of the three models to use for score reporting purposes depends on 1) differences in ability between new and old form groups, as measured by a set of common items (anchor test), and 2) whether the new and old forms are equally reliable, which is typically interpreted to mean of equal test length. These models are based on univariate selection sampling theory. Scores on the relevant selection attribute (the attribute on which the equating samples vary) are assumed to be collinear with scores on the anchor test in the case of the Tucker model and with true scores on both the anchor test and the test form in the case of the Levine models. Scores on the common item set (anchor test) are used to estimate performance of the

BEST COPY AVAILABLE

To equate A3 to X2 and Y3
directly using intact form A3 data

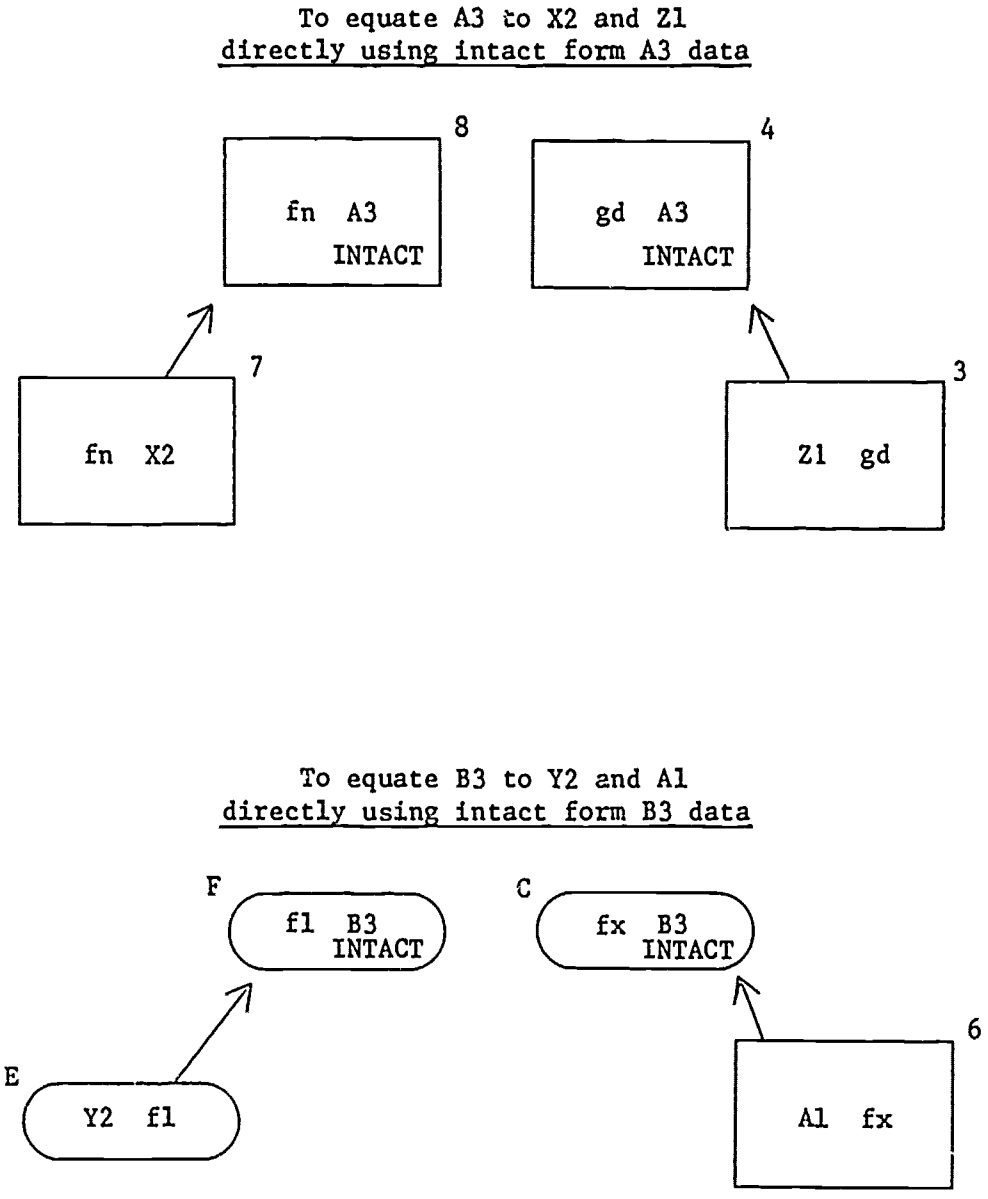


To equate B3 to Y2 and A1
directly using intact form B3 data



BEST COPY

Figure 3: Verbal intact form direct link calibration and linking plan. Upper-case letters followed by one digit designate intact SAT final forms. Lower-case letters designate common item equating sections. Boxes and ovals are numbered to directly correspond to comparable boxes and ovals in Figure 1. Arrows indicate direction in which scaling (linking) took place.



BEST COPY

Figure 4: Mathematics intact form direct link calibration and linking plan. Upper-case letters followed by one digit designate intact SAT final forms. Lower-case letters designate common item equating sections. Boxes and ovals are numbered to directly correspond to comparable boxes and ovals in Figure 2. Arrows indicate direction in which scaling (linking) took place. (In certain instances, a complete LOGIST run contained in Figure 2 is not reproduced here. Only the data from that LOGIST run that is needed for the equating is reproduced.)

combined group of examinees on both the old and new forms of the test, thus simulating by statistical methods the situation in which the same group of examinees takes both forms of the test.

The parameters A and B of the equating transformation are estimated by means of an equation that expresses the relationship between raw scores on two test forms in standard score terms:

$$\frac{x - M_x}{S_x} = \frac{y - M_y}{S_y} \quad , \tag{3}$$

where x and y refer to the test scores to be equated, and M and S refer to the means and standard deviations of the scores in some group of examinees. Methods using the above equation differ in their identification of the means and standard deviations to be estimated. The Tucker and Levine Equally Reliable methods are based on the estimated means and standard deviations of observed scores whereas the Levine Unequally Reliable method is based on the estimated means and standard deviations of true scores.

The formulas for computing the A and B parameters for the Tucker, Levine Equally Reliable, and Levine Unequally Reliable models are given in Table 5. As noted in Table 5, the formulas for the Levine models require error variance estimates. Angoff's method (1953) of estimating error variances is used for operational linear equating. This method assumes that the test to be equated and the anchor test are parallel except for length.

When a new form is equated to two old forms, the final linear parameters to put the new form on scale are arrived at in the following fashion. Each of the old forms has linear parameters for placing it on scale; these parameters are combined with linear parameters generated from the equating relationship to derive parameters to put the new form on scale. There will

Table 5

Formulas for Linear Conversion Parameters

Tucker

$$A = (S^2_{yb} + C^2_{yvb}(S^2_{vc} - S^2_{vb})/S^4_{vb})^{\frac{1}{2}}(S^2_{xa} + C^2_{xva}(S^2_{vc} - S^2_{va})/S^4_{va})^{-\frac{1}{2}}$$

$$B = M_{yb} + C_{yvb}(M_{vc} - M_{vb})/S^2_{vb} - AM_{xa} - AC_{xva}(M_{vc} - M_{va})/S^2_{va}$$

Levine Equally Reliable

$$A = (S^2_{yb} + (S^2_{yb} - S^2_{y''b})(S^2_{vc} - S^2_{vb})/(S^2_{vb} - S^2_{v''b}))^{\frac{1}{2}}$$
$$(S^2_{xa} + (S^2_{xa} - S^2_{x''a})(S^2_{vc} - S^2_{va})/(S^2_{va} - S^2_{v''a}))^{-\frac{1}{2}}$$

$$B = M_{yb} + (M_{vc} - M_{vb})((S^2_{yb} - S^2_{y''b})/(S^2_{vb} - S^2_{v''b}))^{\frac{1}{2}}$$
$$- AM_{xa} - A(M_{vc} - M_{va})((S^2_{xa} - S^2_{x''a})/(S^2_{va} - S^2_{v''a}))^{\frac{1}{2}}$$

Levine Unequally Reliable

$$A = ((S^2_{yb} - S^2_{y''b})/(S^2_{vb} - S^2_{v''b}))^{\frac{1}{2}}((S^2_{xa} - S^2_{x''a})/(S^2_{va} - S^2_{v''a}))^{-\frac{1}{2}}$$

$$B = M_{yb} + (M_{vc} - M_{vb})((S^2_{yb} - S^2_{y''b})/(S^2_{vb} - S^2_{v''b}))^{\frac{1}{2}} - AM_{xa}$$

Angoff Error Variance Estimates (Anchor Test External to Total Test)

$$S^2_{p''g} = (S^2_{pg}S^2_{vg} - C^2_{pvg})/(S^2_{vg} + C_{pvg})$$

$$S^2_{v''g} = (S^2_{pg}S^2_{vg} - C^2_{pvg})/(S^2_{pg} + C_{pvg})$$

Notation

New Test Form	X
Old Test Form	Y
Either New or Old Test Form	P
Anchor Test	V
Observed Score	x, y, v, p
Error Score	x", y", v", p"
Group Taking Test X and Test V	a
Group Taking Test Y and Test V	b
Group Taking Test P and Test V	g
Combined Group	c or (a + b)
Mean	M
Standard Deviation	S
Covariance	C

BEST COPY

be a set of parameters for each equating to each old form; the final set of parameters are arrived at by averaging the parameters from each of the single equatings.

Although there are a number of equating techniques possible when using IRT, this study was concerned only with true formula score equating (Lord, 1980). The expected value of an examinee's observed formula score is defined as his or her true formula score. For the true formula score, ξ, we have

$$\xi = \sum_{i=1}^{n} \left[\frac{(k_i + 1)}{k_i} P_i(\theta) - \frac{1}{k_i} \right] \qquad (4)$$

where n is the number of items in the test and (k_i+1) is the number of choices for item i. If we have two tests measuring the same ability θ, then true formula scores ξ and η from the two tests are related by the equations

$$\begin{aligned} \xi &= \sum_{i=1}^{n} \left[\frac{(k_i + 1)}{k_i} P_i(\theta) - \frac{1}{k_i} \right] \\ \eta &= \sum_{j=1}^{m} \left[\frac{(k_j + 1)}{k_j} P_j(\theta) - \frac{1}{k_j} \right] \end{aligned} \qquad (5)$$

Clearly, for a particular θ corresponding true scores ξ and η have identical meaning. They are said to be equated.

Because true formula scores below the chance score level are undefined for the three-parameter logistic model, some method must be established to obtain a relationship between scores below the chance level on the two test forms to be equated. The approach used for this study (Lord, 1980) was to estimate the mean (M) and standard deviation (S) of below chance level scores on the two tests to be equated via the formulas

BEST COPY

$$M = \sum_{i=1}^{n} \left[c_i (c_i + 1)/k_i - 1/k_i \right] , \quad \text{and} \qquad (6)$$

$$S^2 = \sum_{i=1}^{n} (c_i - c_i^2) (k_i + 1)^2/k_i^2 ,$$

where n is the number of items in the test, (k_i+1) is the number of choices for item i, and c_i is the psuedo-guessing parameter for item i; and then to use these estimates to do a simple linear equating (see equation (3)) between the two sets of below chance level scores.

In practice, true score equating is carried out by substituting estimated parameters into the equations (5) and (6). Paired values of ξ and η are then computed for a series of arbitrary values of θ. Since we cannot know an examinee's true formula score, we act as if relationships (5) and (6) apply to an examinee's observed formula score.

Two further points require clarification. First, the mechanics of doing IRT true-score equating based on pretest data (pre-equating) and based on intact final form data are exactly the same. What differs are the item parameter estimates that are used to calculate $P_i(\theta)$ in equation (4). In one instance the parameters have been calibrated for the item when given in a pretest, and in the other instance, when the item was given as part of an intact final form. Second, when performing score equating to two old forms using IRT true-score equating techniques, a conversion table is generated for each new form-old form relationship and then the corresponding entries in each table are simply averaged to generate the final table.

BEST COPY

Results

SAT-verbal and SAT-mathematical Pre-equating Plots and Tables

A number of figures and tables have been prepared to summarize the results of this study. Because the verbal and mathematical equatings for 3ASA3 and for 3BSA3 are independent and meant to serve as replications of the pre-equating process, the figures for the verbal and mathematical sections of each form can be viewed separately. The verbal and mathematical sections of each of the forms were equated to two old forms; thus, there are figures for each of the single equatings and then the equating resulting from the averaging of the single equatings. Figures 5-7 contain the equating results for 3ASA3 verbal, while Figure 8-10 contain comparable results for 3BSA3 verbal. Results from the equating of 3ASA3 mathematical are contained in Figures 11-13, while comparable results for 3BSA3 mathematical are contained in Figures 14-16. Tables 1-12 in the Appendix give point by point conversions for each of the equatings performed for Forms 3ASA3 and 3BSA3 of SAT-verbal and SAT-mathematical.

In the figures for each equating performed, there are two plots. The first plot compares the raw to scaled score conversion line resulting from the IRT pre-equating to one of the three comparison conversion lines, resulting from either the intact form calibration system IRT equating, the intact form direct link IRT equating, or the intact form linear equating actually used operationally for score reporting purposes. The second plot contains residuals. These residuals are simple differences between scaled scores resulting from the IRT pre-equating and one of the comparison equatings for each possible formula score point. The plots use the

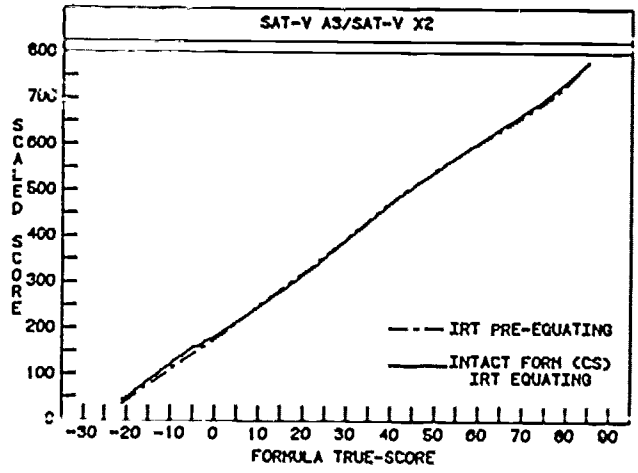
BEST COPY AVAILABLE

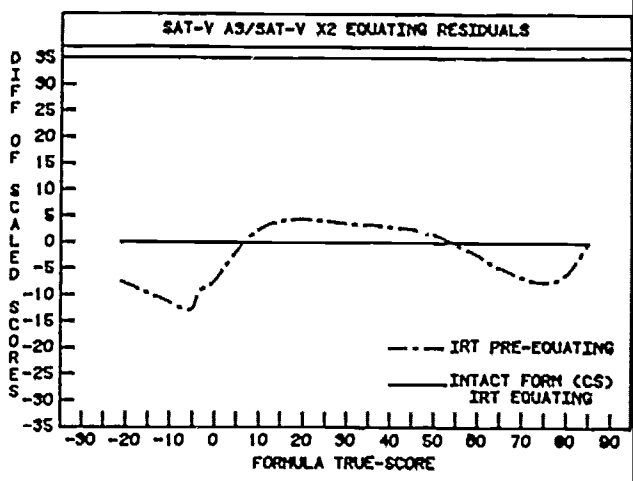
Equating Plot

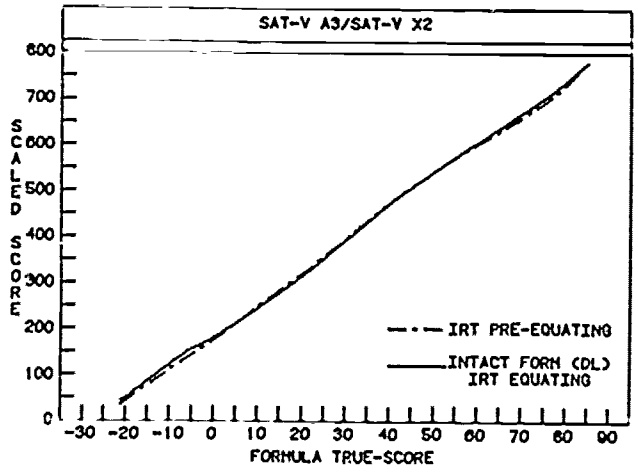
Residual Plot

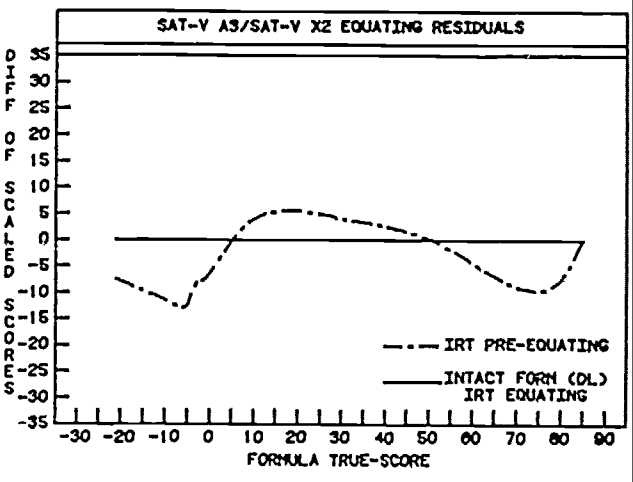
Intact Form Direct Link IRT Equating

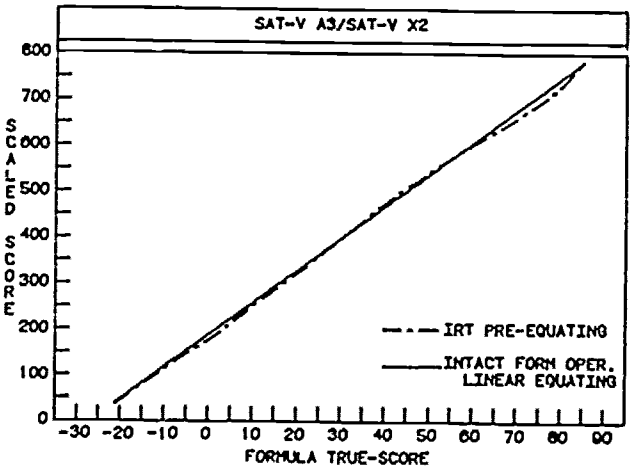
Intact Form Operational Linear Equating











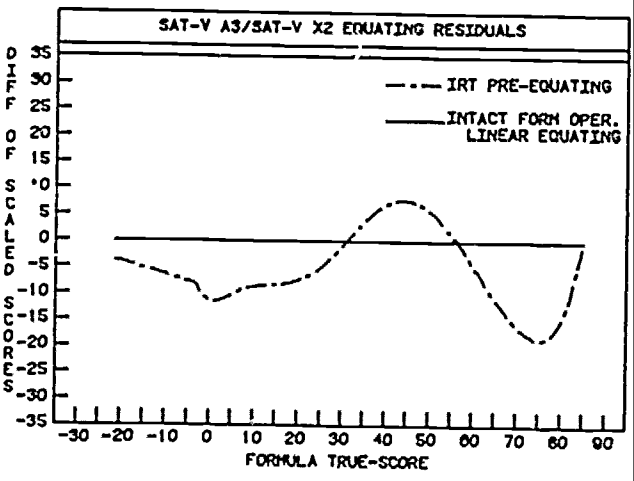


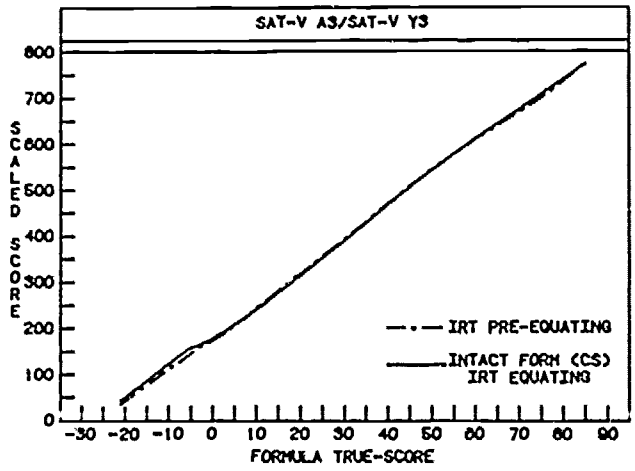
Figure 5: SAT-verbal Form 3ASA3 equated to SAT-verbal Form XSA2 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

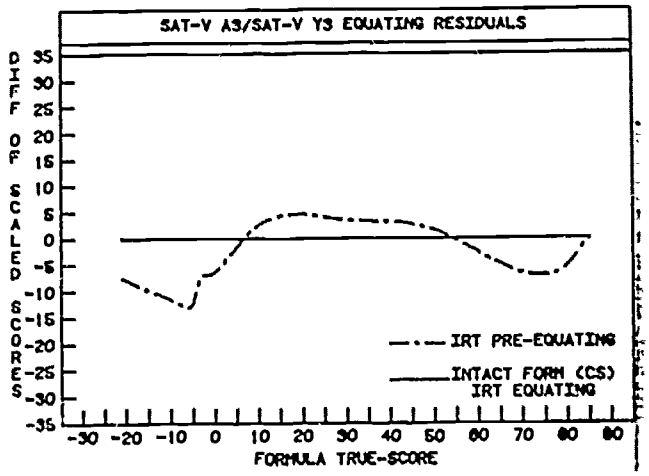
BEST COPY

Equating Plot

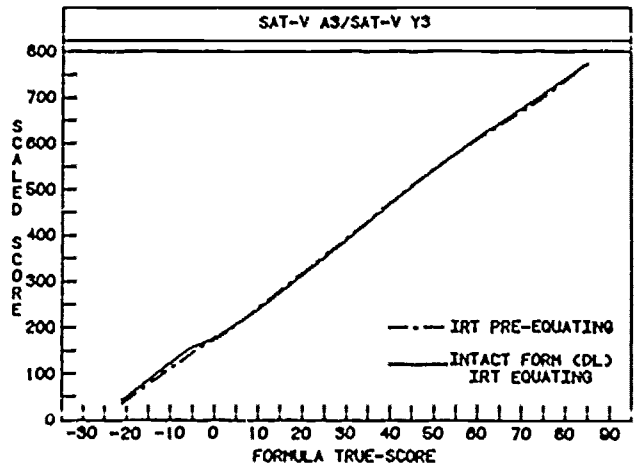
Residual Plot

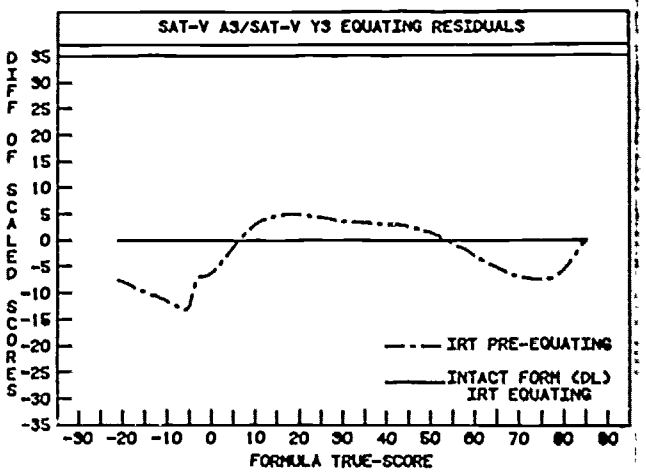
Intact Form Calibration System IRT Equating



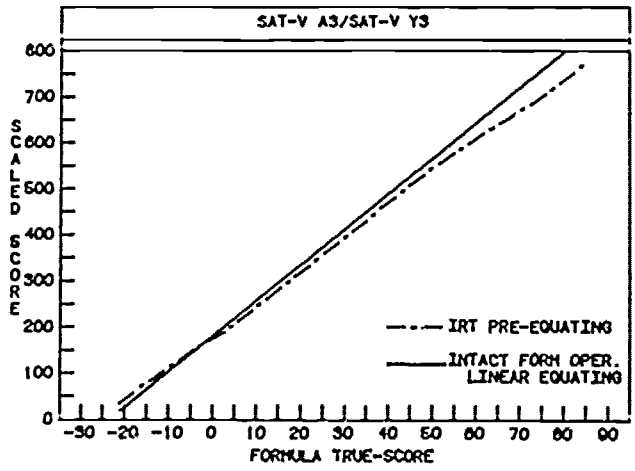


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



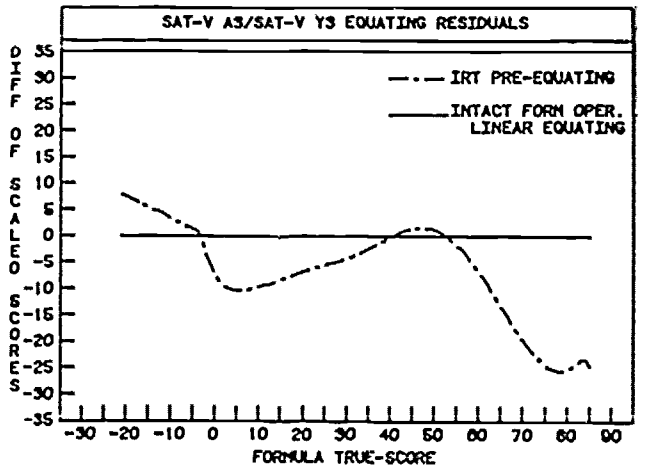


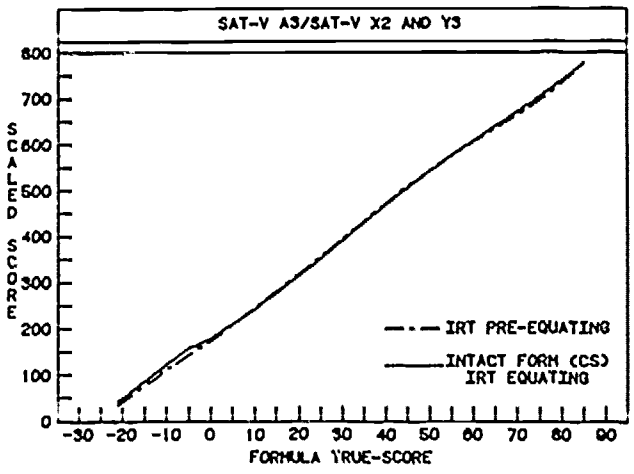
Figure 6: SAT-verbal Form 3ASA3 equated to SAT-verbal Form YSA3 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

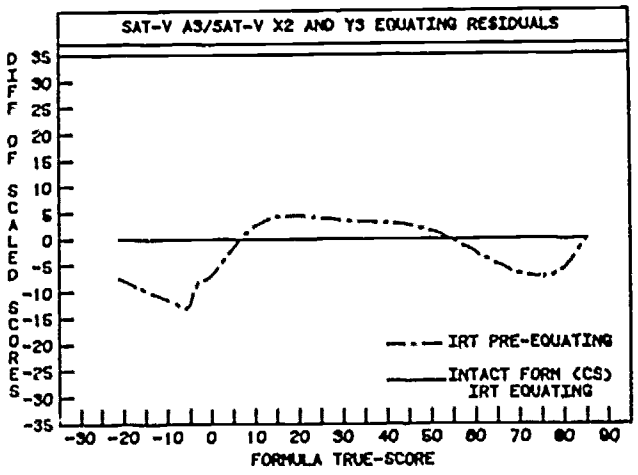
BEST COPY

Equating Plot

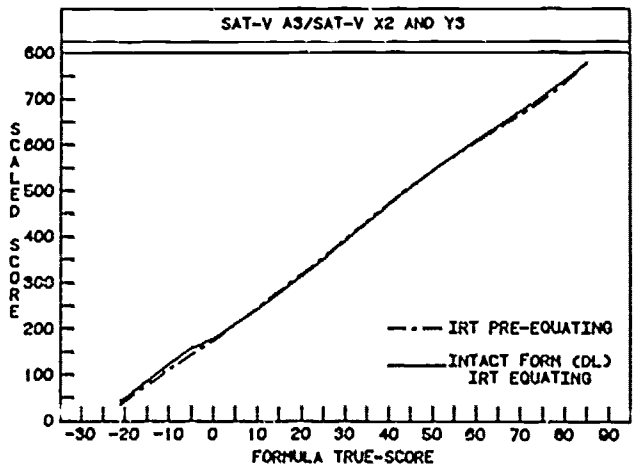
Residual Plot

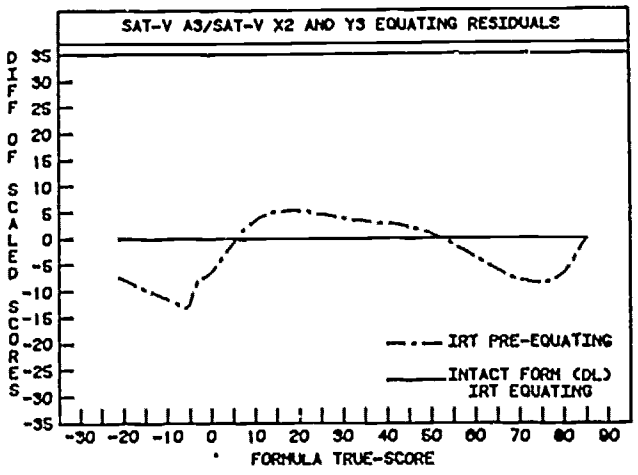
Intact-Form Calibration System IRT Equating



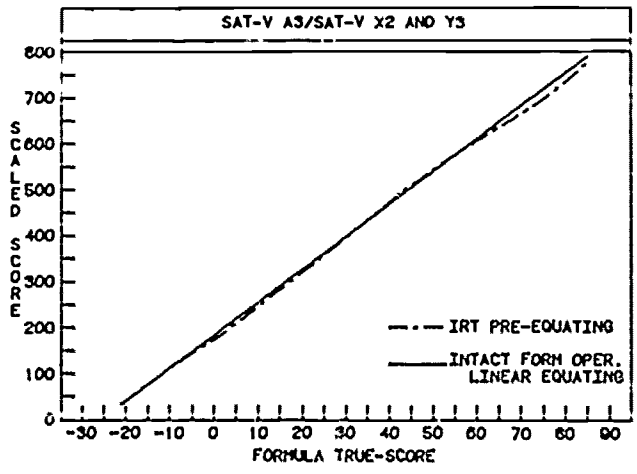


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



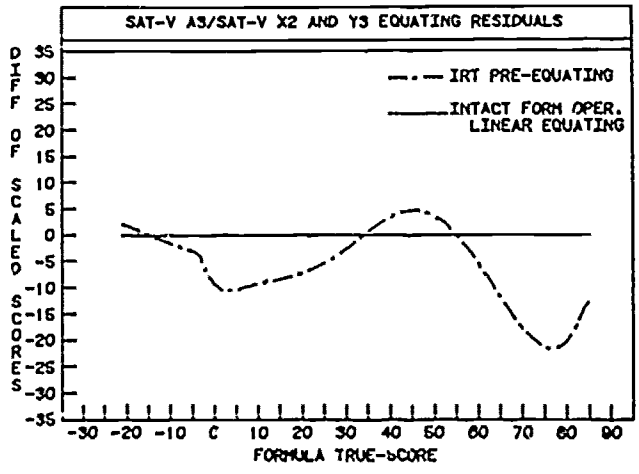
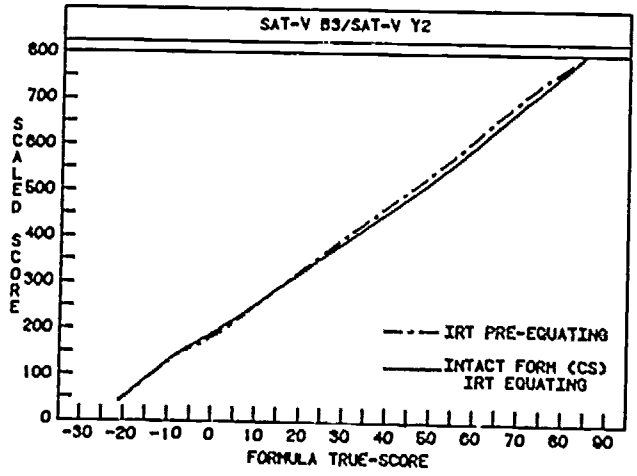


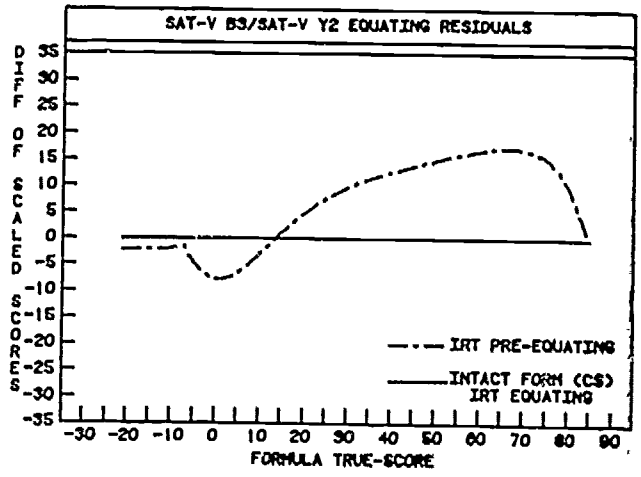
Figure 7: SAT-verbal Form 3ASA3 equated to SAT-verbal Form XSA2 and SAT-verbal Form YSA3 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

BEST COPY

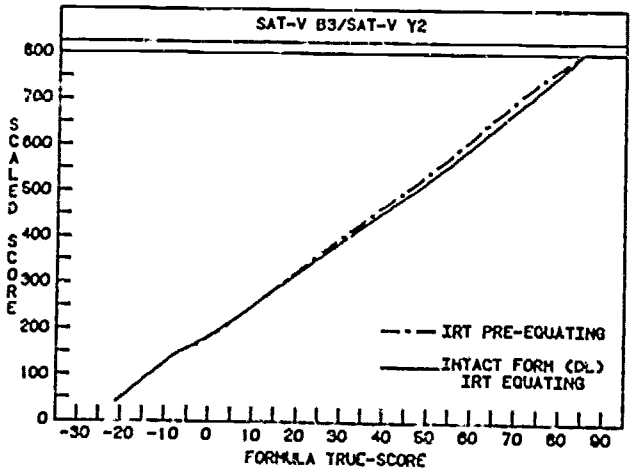
Equating Plot | Residual Plot

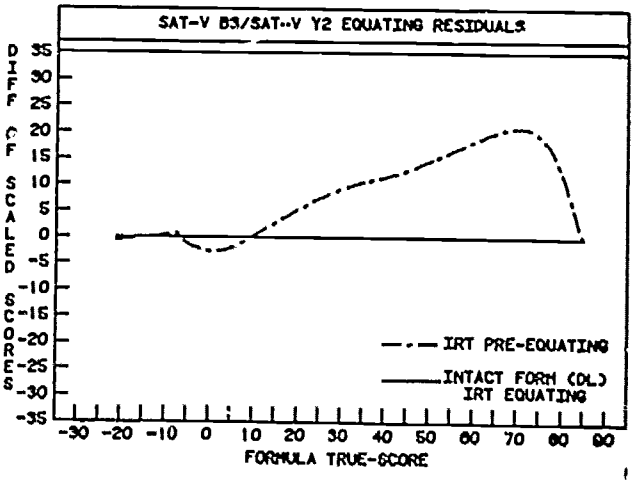
Intact Form Calibration System IRT Equating



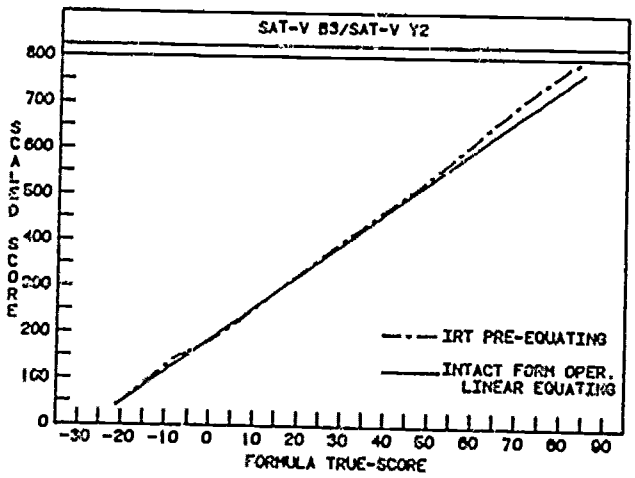


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



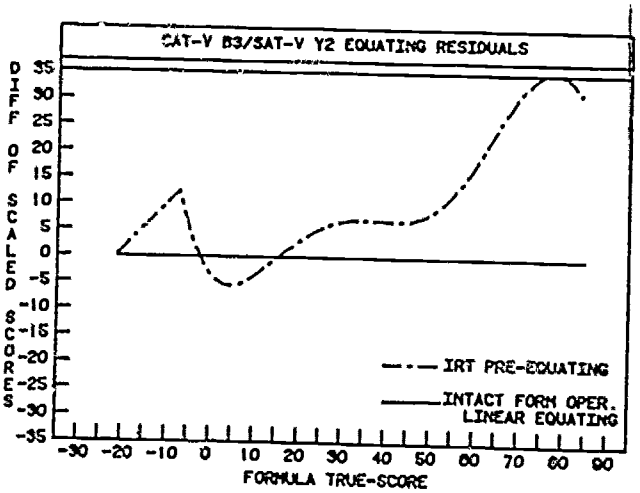


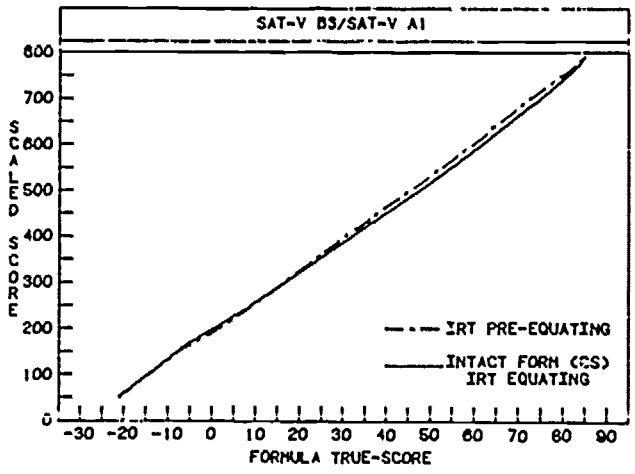
Figure 8: SAT-verbal Form 3BSA3 equated to SAT-verbal form YSA2 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

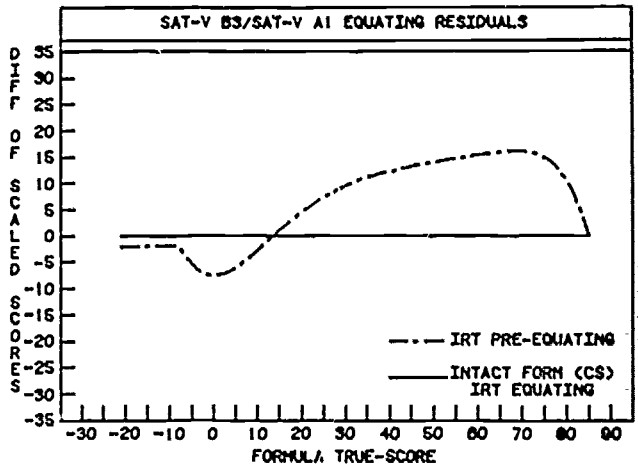
BEST COPY AVAILABLE

Equating Plot

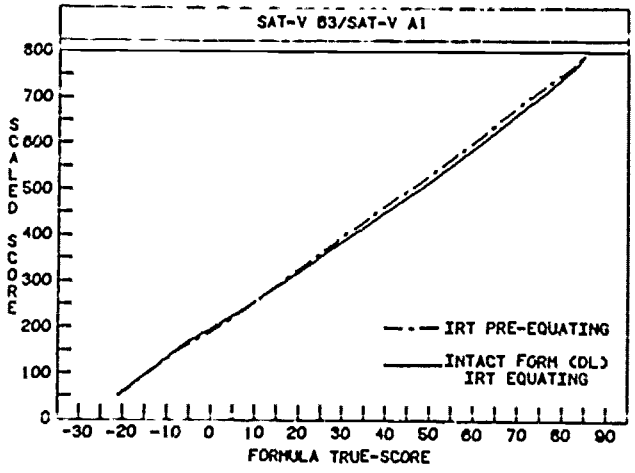
Residual Plot

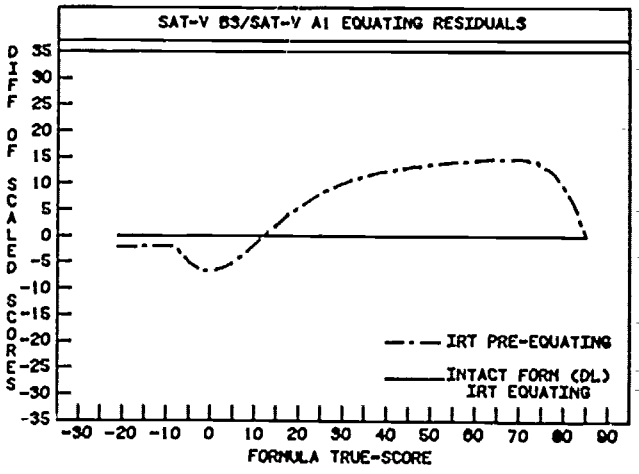
Intact Form Calibration System IRT Equating



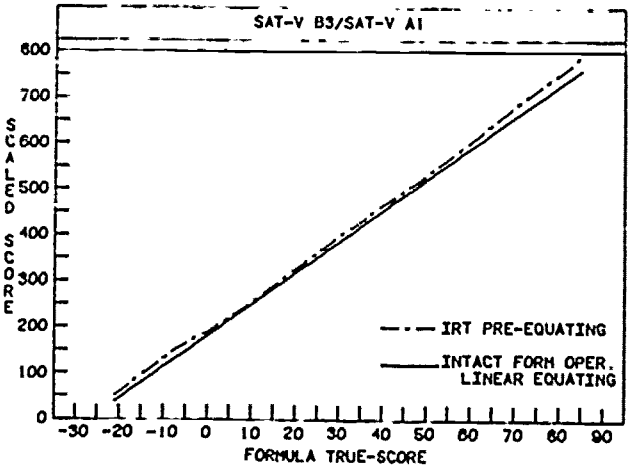


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



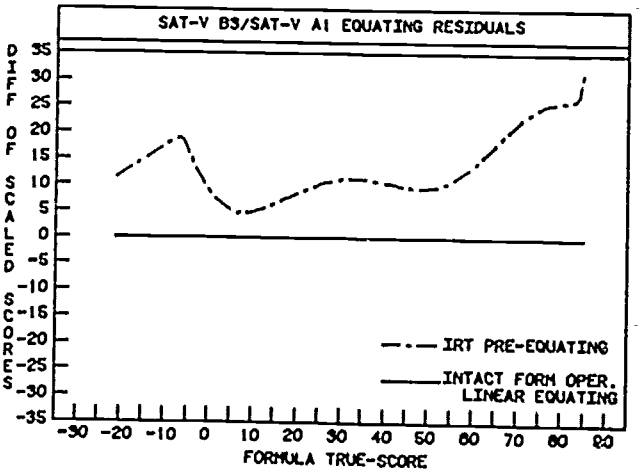
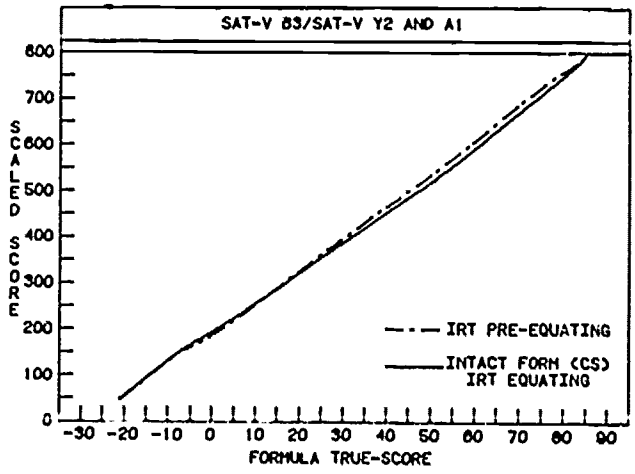


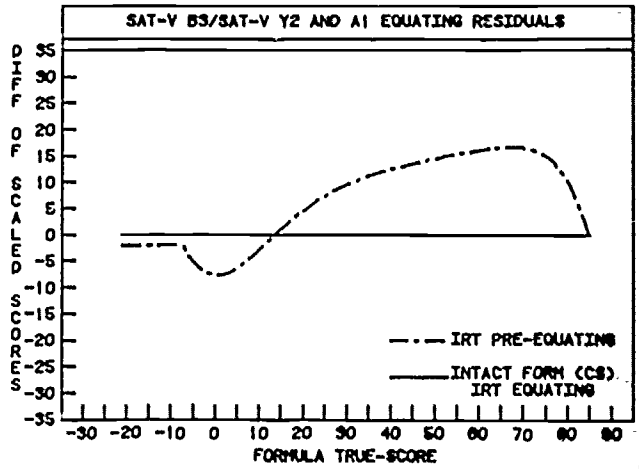
Figure 9: SAT-verbal Form 3BSA3 equated to SAT-verbal Form 3ASA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

BEST COPY

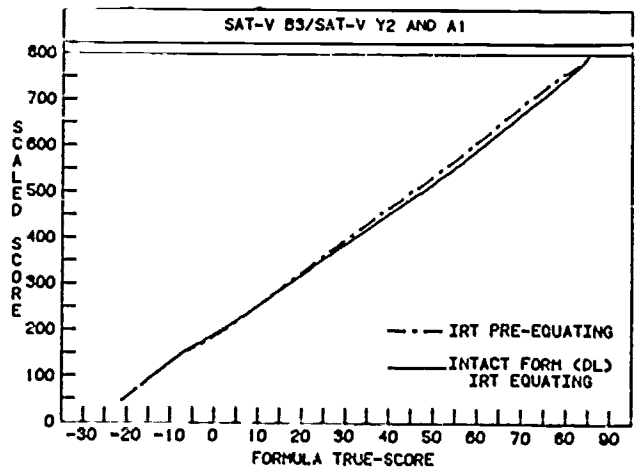
Equating Plot | Residual Plot

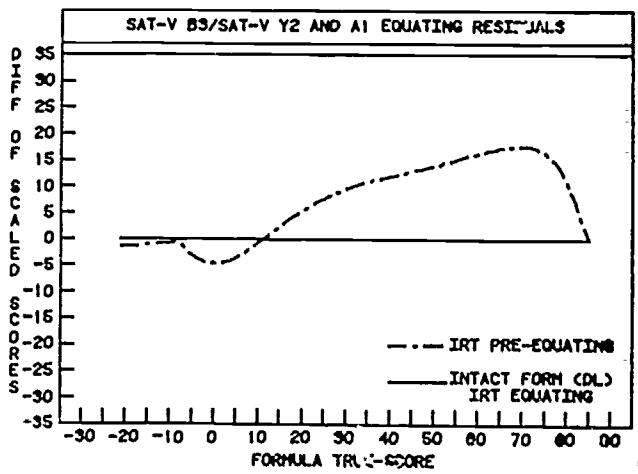
Intact Form Calibration System IRT Equating



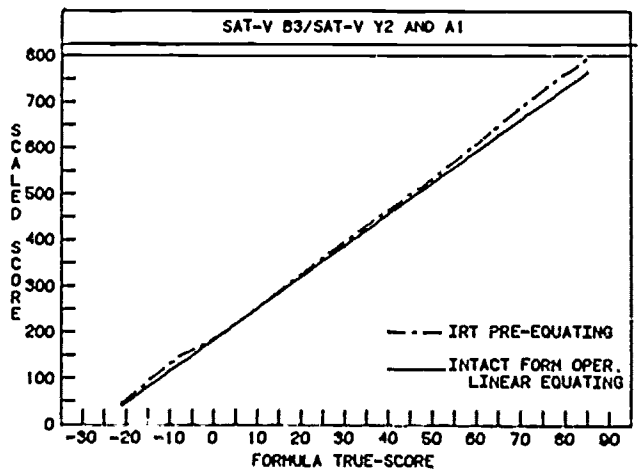


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



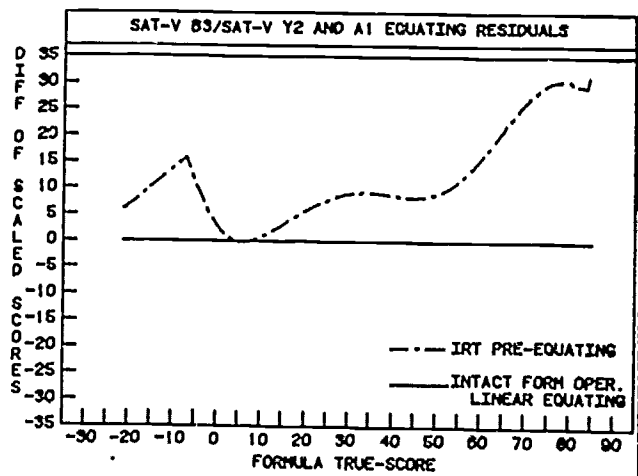
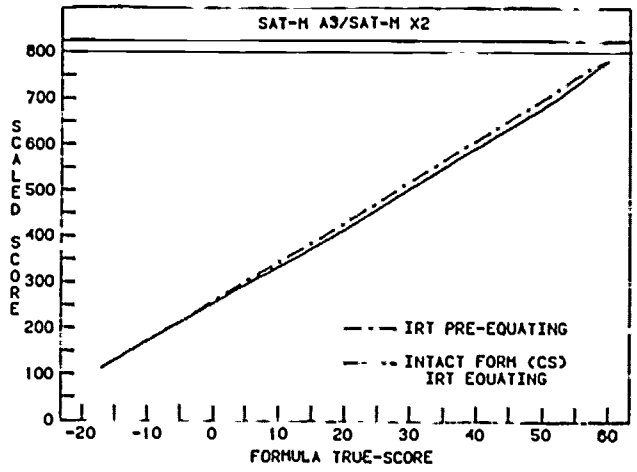


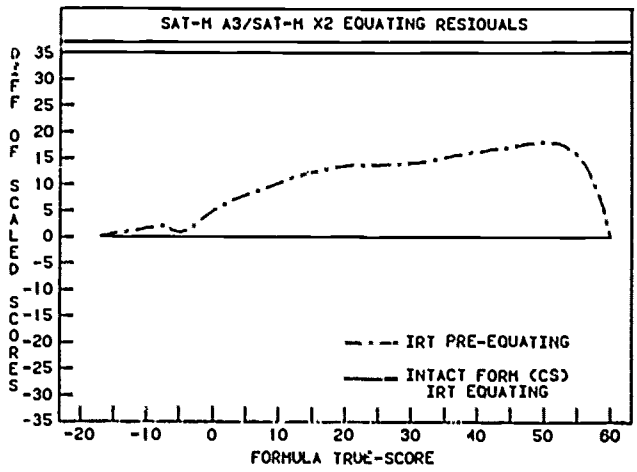
Figure 10: SAT-verbal Form 3BSA3 equated to SAT-verbal Form YSA2 and SAT-verbal Form 3ASA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

Equating Plot

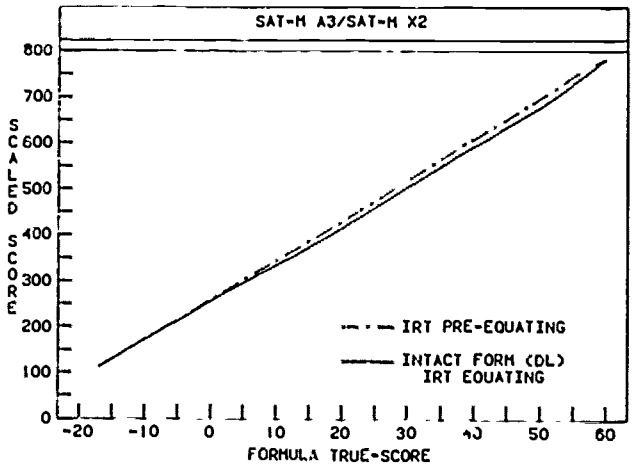
Residual Plot

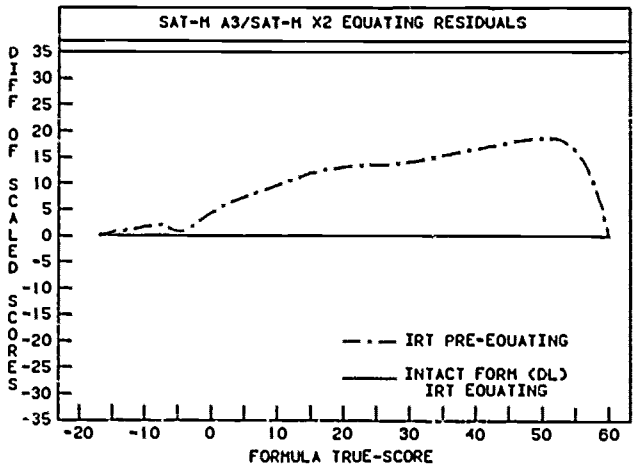
Intact Form Calibration System IRT Equating



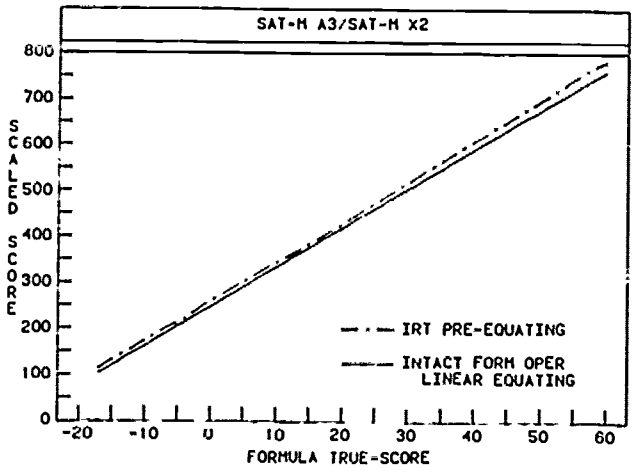


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



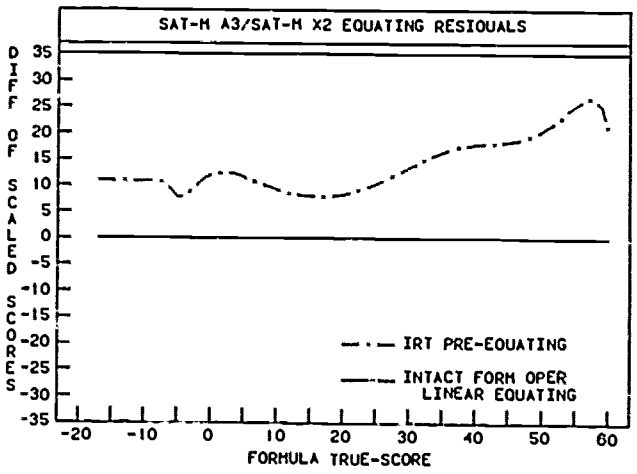


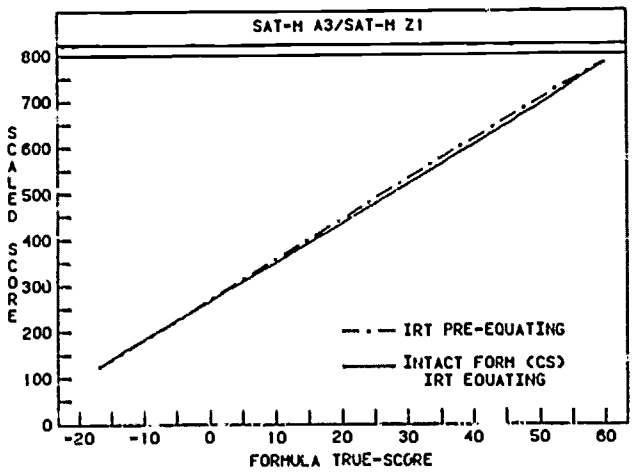
Figure 11: SAT-math Form 3ASA3 equated to SAT-math Form XSA2 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact from calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

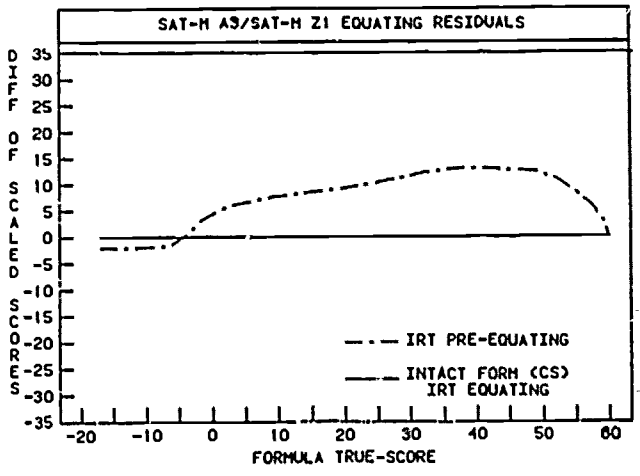
BEST COPY

Equating Plot

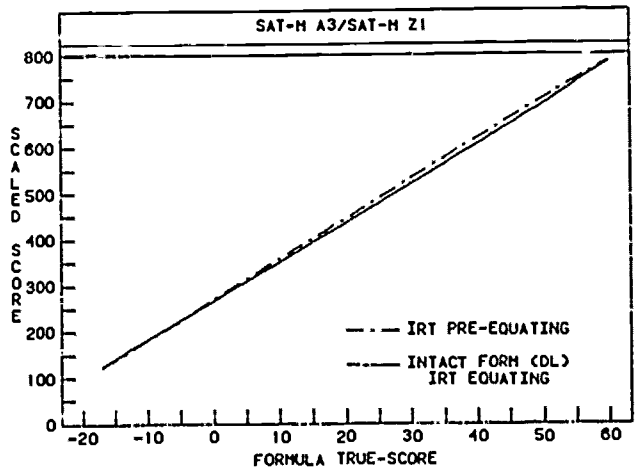
Residual Plot

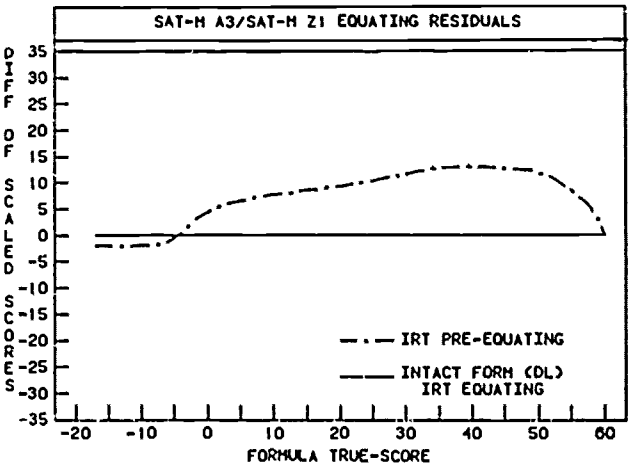
Intact Form Calibration System IRT Equating



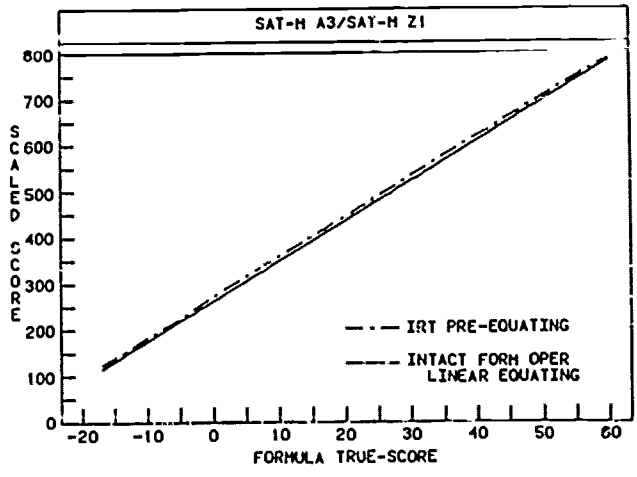


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



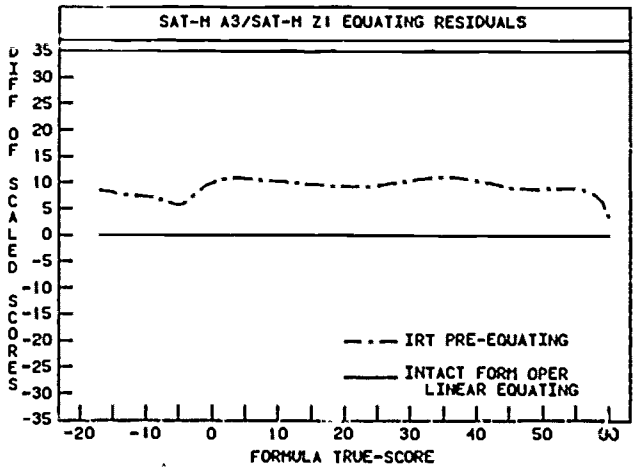


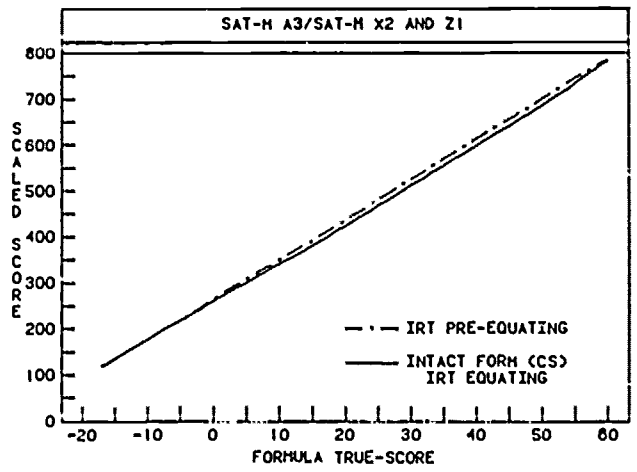
Figure 12: SAT-math form 3ASA3 equated to SAT-math Form ZSA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformation, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

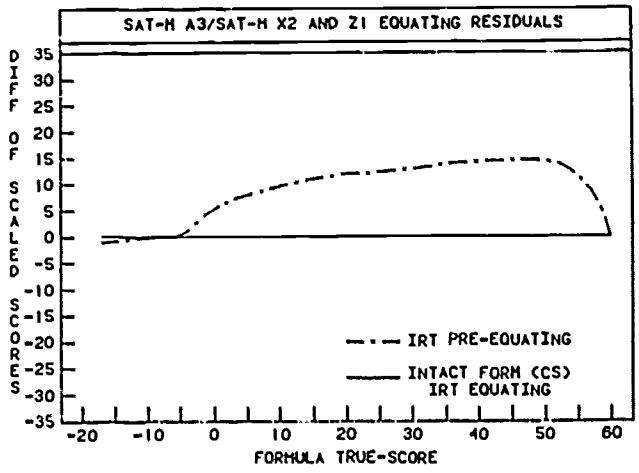
BEST COPY AVAILABLE

Equating Plot

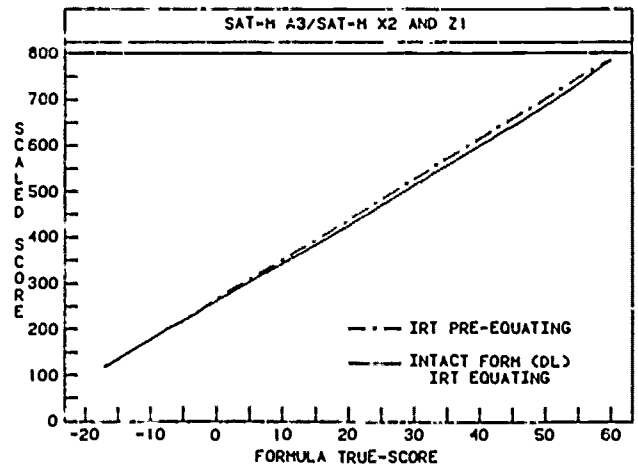
Residual Plot

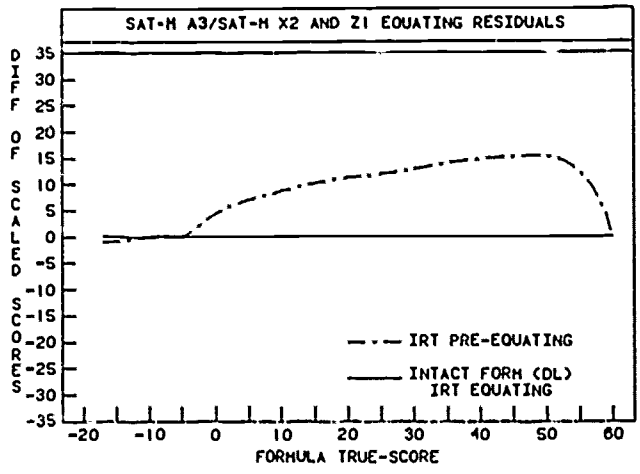
Intact Form Calibration System IRT Equating



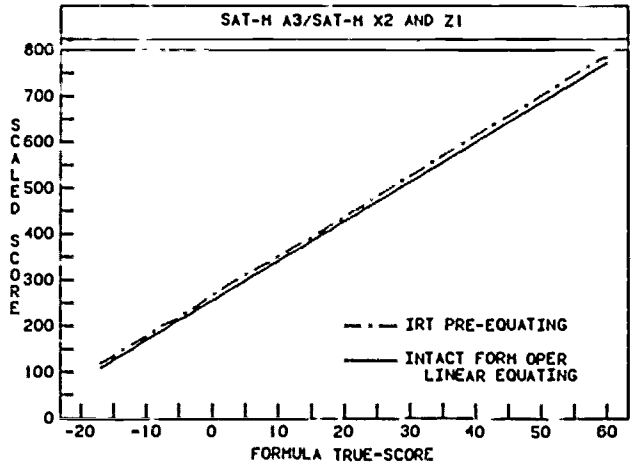


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



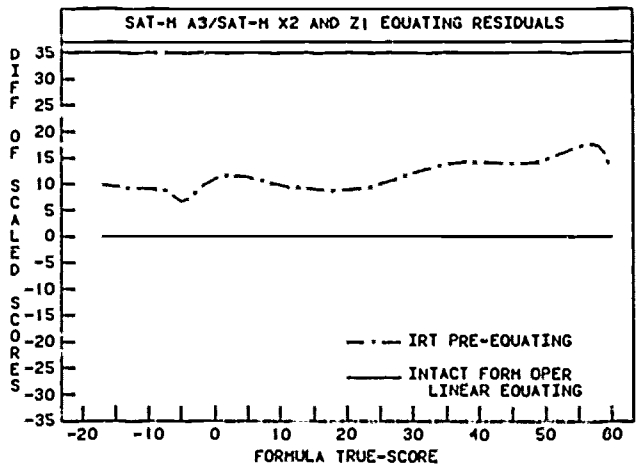
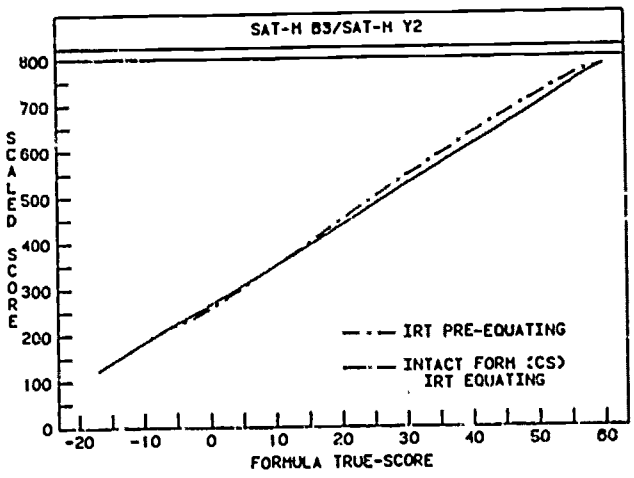


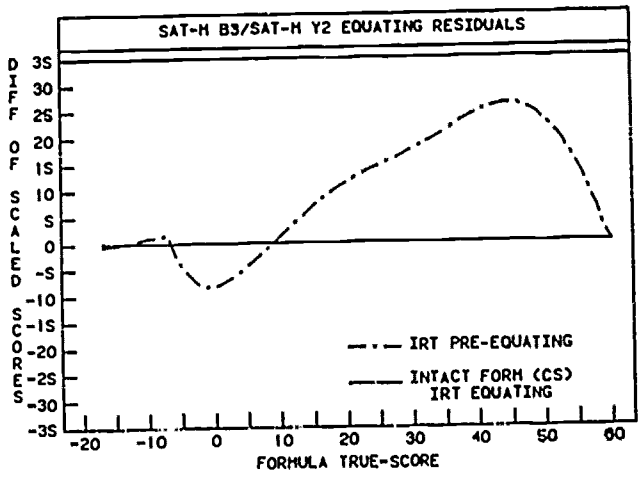
Figure 13: SAT-math Form ZSA3 equated to SAT-math Form XSA2 and SAT-math Form ZSA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

BEST COPY

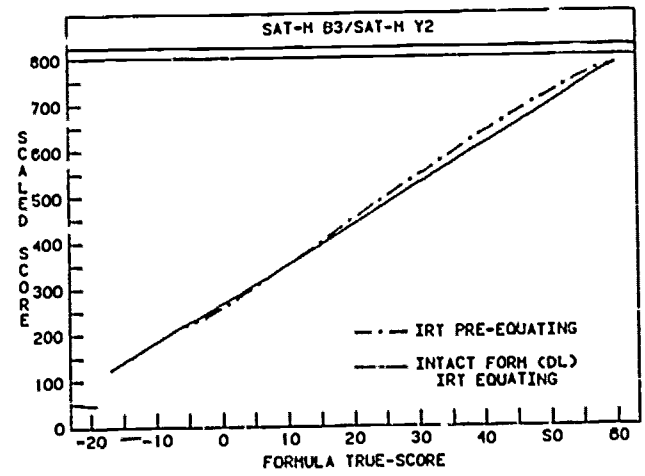
Equating Plot | Residual Plot

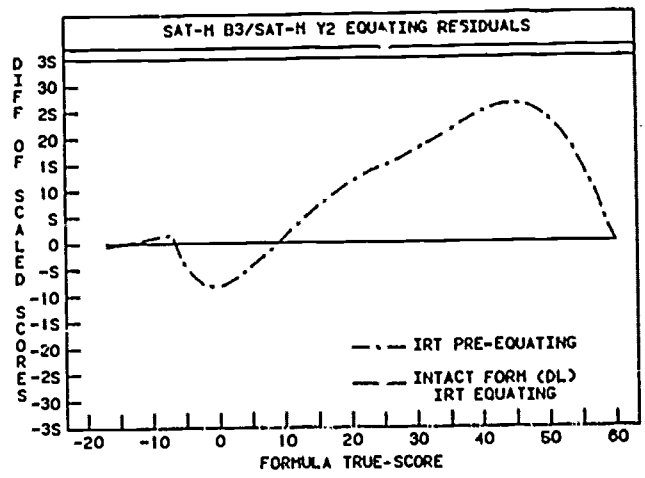
Intact Form Calibration System IRT Equating



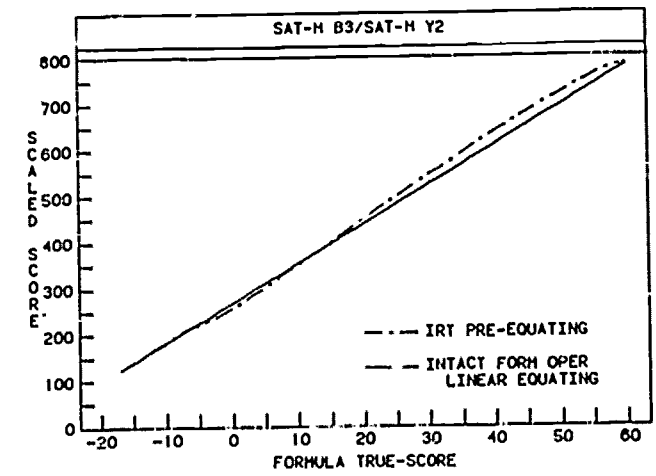


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



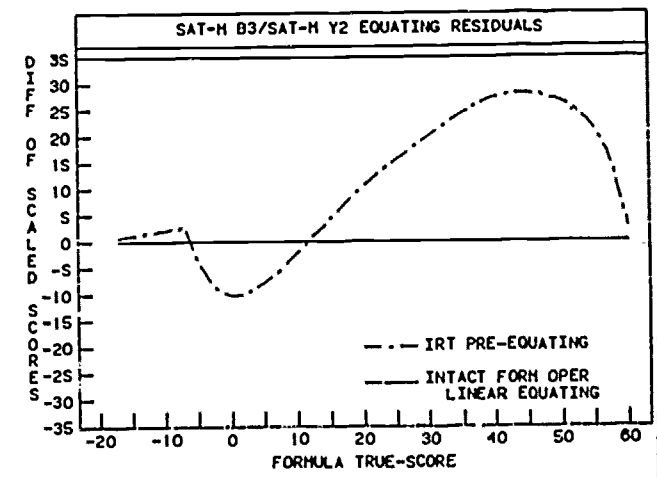


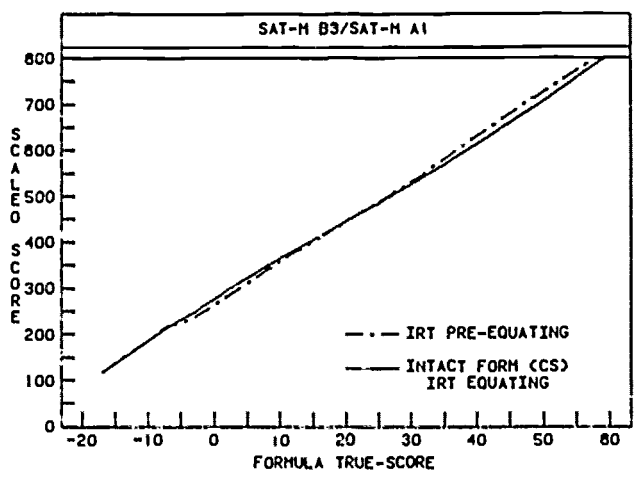
Figure 14: SAT-math Form 3BSA3 equated to SAT-math Form YSA2 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

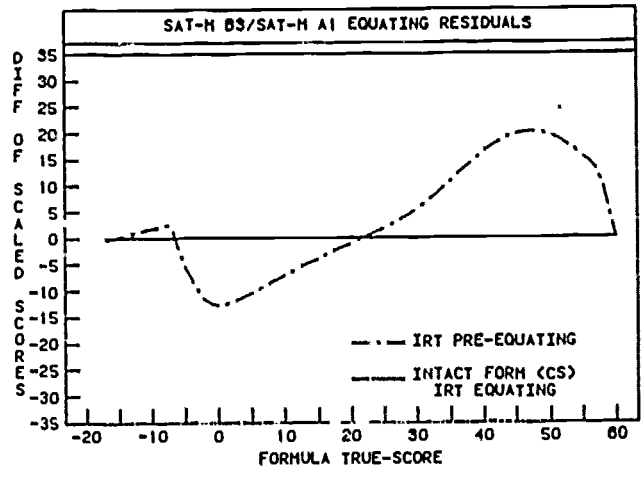
BEST COPY AVAILABLE

Equating Plot

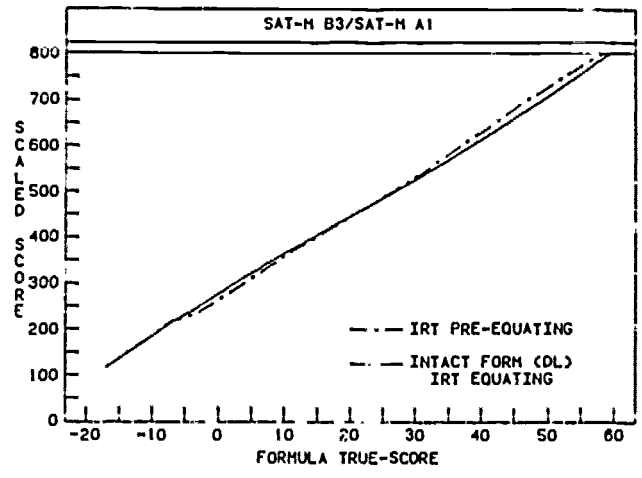
Residual Plot

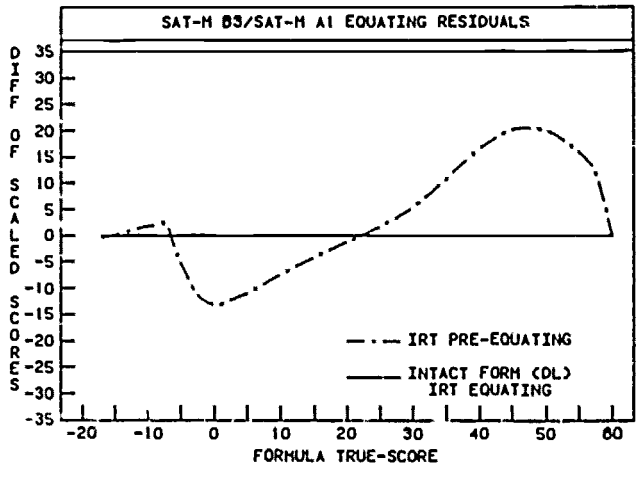
Intact Form Calibration System IRT Equating



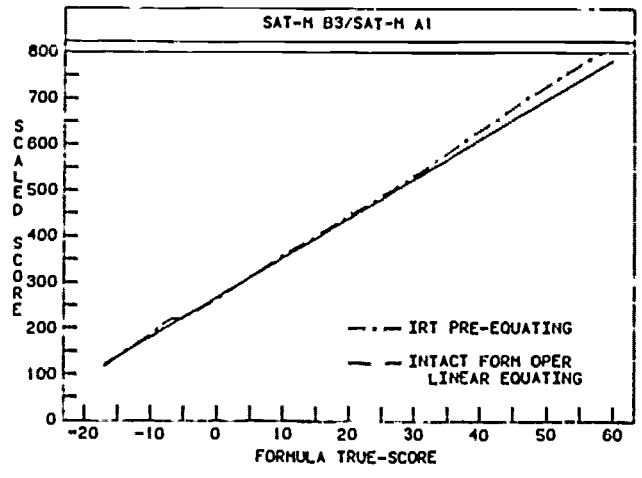


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



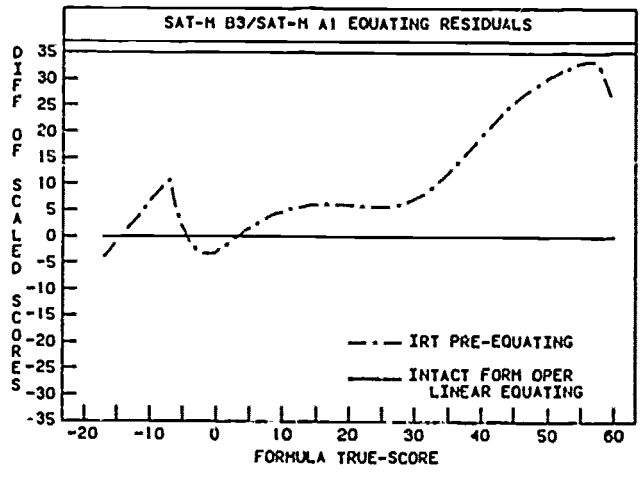


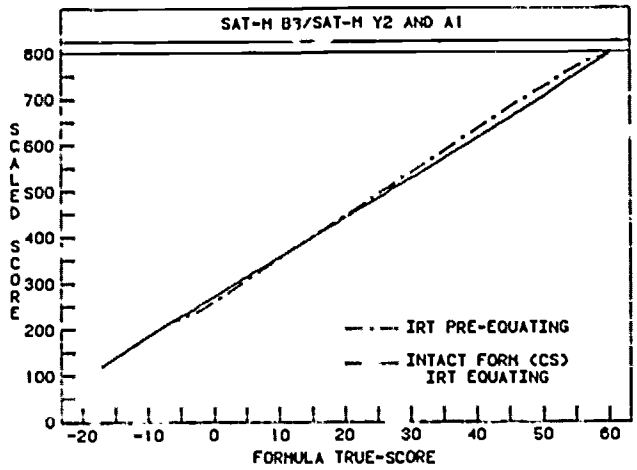
Figure 15: SAT-math Form 3BSA3 equated to SAT-math Form 3ASA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, and 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

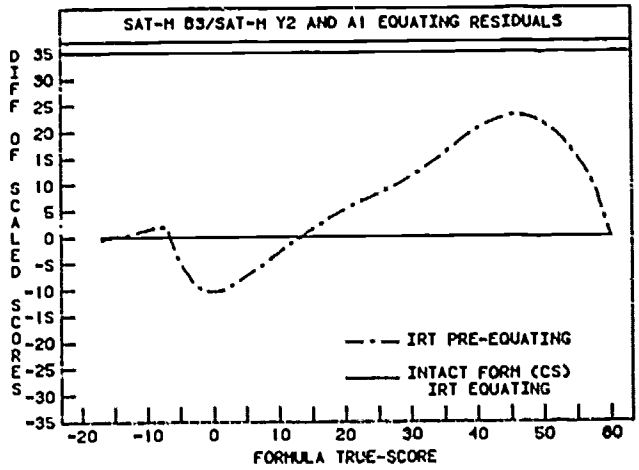
BEST COPY AVAILABLE

Equating Plots

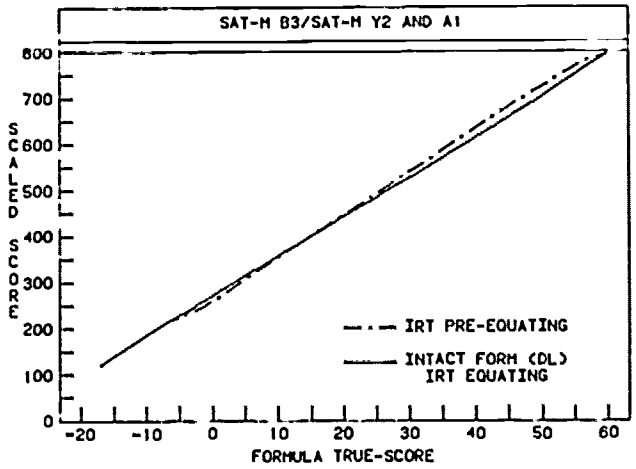
Residual Plot

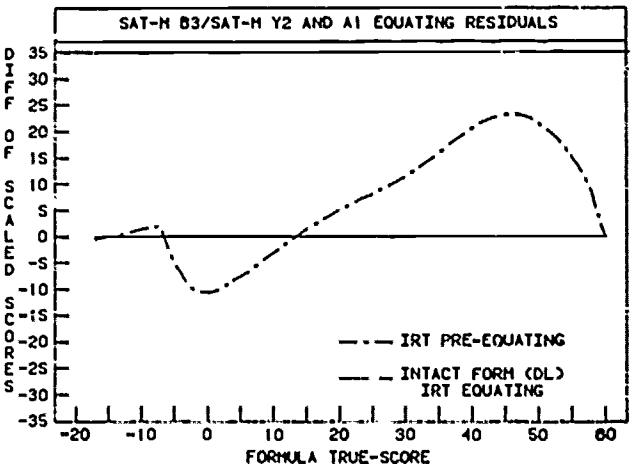
Intact Form Calibration System IRT Equating



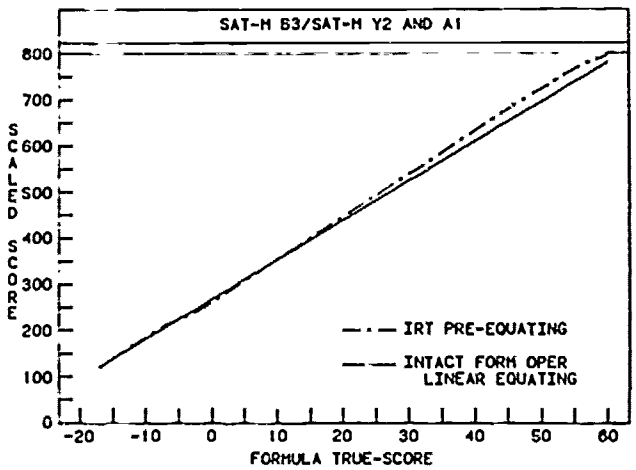


Intact Form Direct Link IRT Equating





Intact Form Operational Linear Equating



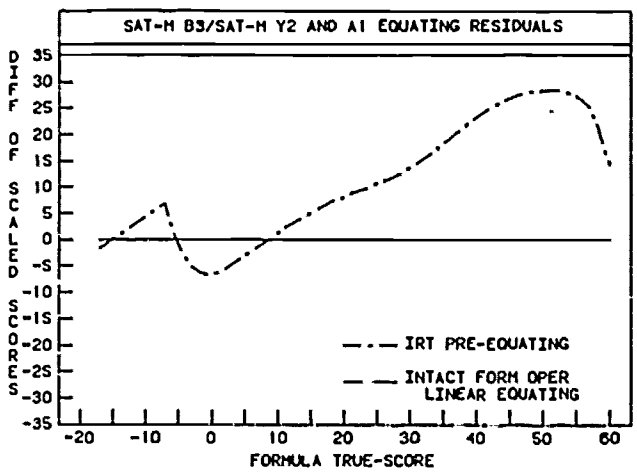


Figure 16: SAT-math Form 3BSA3 equated to SAT-math Form YSA2 and SAT-math Form 3ASA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to corresponding intact form calibration system IRT, direct link IRT, and operational linear equating raw to scaled score transformations, 2) differences between scaled scores (IRT pre-equating - comparison equating) resulting from the equatings.

comparison equating (intact form calibration IRT equating, intact form direct link IRT equating, or intact form linear equating) as the baseline and show differences between the pre-equating results and the baseline equating results across the formula score scale. As mentioned earlier, the intact form calibration system and direct link IRT equatings were chosen as baseline equatings for these residual plots because these sorts of IRT equatings, in particular the direct link IRT equating, have been shown in previous studies to be viable equating methods for SAT data, and provide good criterion equatings against which to evaluate the pre-equating results. The intact form linear equating was also used as a baseline because this was the method actually used to put the 3ASA3 and 3BSA3 verbal and mathematical sections on scale operationally. Of the three comparison equatings, the intact form direct link equating should provide the best criterion against which to evaluate the pre-equating results in that 1) the relationship between the parameter estimates, for the forms to be equated, from the separate LOGIST runs have been influenced by no more than one intervening scaling, unlike the case with the intact form calibration system equating, and 2) in contrast with linear equating, curvilinear relationships are permitted. The residual plots, in conjunction with data presented in Tables 6 and 7, to be described next, provide much of the data upon which to evaluate the pre-equating results of this study.

Table 6 provides the verbal scaled score means and standard deviations for Forms 3ASA3 and 3BSA3 that would have resulted from use for score reporting purposes of pre-equating, intact form calibration system IRT equating, intact form direct link IRT equating, and intact form linear equating to the old forms. Table 7 provides comparable data for 3ASA3 and

Table 6

Scaled Score Summary Statistics Resulting from Application of Four Equating Methods
SAT-verbal Sections of Forms 3ASA3 and 3BSA3

Form	N		IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
3ASA3	126,788	M	437.04	437.01	441.45	439.26
		S.D.	111.91	111.30	108.34	109.65
3BSA3	253,354	M	430.25	430.42	431.42	440.39
		S.D.	105.99	105.57	106.53	110.55

Table 7

Scaled Score Summary Statistics Resulting from Application of Four Equating Methods
SAT-math Sections of Forms 3ASA3 and 3BSA3

Form	N		IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
3ASA3	126,788	M	484.77	484.78	485.18	496.65
		S.D.	112.94	113.40	113.37	115.27
3BSA3	253,354	M	481.20	481.02	477.80	489.06
		S.D.	112.82	113.09	112.85	121.58

3BSA3 mathematical. The means and standard deviations were computed using frequencies for the total groups taking Forms 3ASA3 and 3BSA3 at the respective initial intact form administrations.

SAT-verbal Pre-equating Results

Based on the data presented for the verbal section of Form 3ASA3, it is clear that the pre-equating was reasonably successful. In no residual plot is the difference between the scaled score resulting from the pre-equating and the comparison intact form calibration system IRT or direct link IRT equatings more than 15 score points on a scale containing 600 score points. The differences between the pre-equating results and the intact form linear results are greater than the differences resulting from the intact form IRT equatings, particularly at the upper end of the formula score scale. This is because all three IRT equatings demonstrate that the raw to scaled score conversion is curvilinear in this region, and the linear equating cannot account for this curvilinearity. The differences in scaled score means and standard deviations presented in Table 6 are very small. The scaled score means and standard deviations resulting from the two IRT methods used as criteria are almost identical. The scaled score mean resulting from the pre-equating lies between the scaled score mean resulting from either the intact form calibration system or direct link IRT equatings and the scaled score mean resulting from the intact form linear equating. The scaled score difference between the mean resulting from the pre-equating and any of the other equatings is about 2 points. What is particularly interesting to note is the pattern of the residuals plots for the comparison of the pre-equating results with the intact form calibration system and direct link IRT equating

results, displayed in Figures 5-7. The patterns of residuals are the same across both the single equatings and, as a result, the average equating. The pre-equating results in lower scaled scores at the bottom and top of the formula score scale and slightly higher scaled scores in the middle region. As mentioned earlier, at no point are these differences greater than 15 scaled score points, and hence, although the pattern of differences is consistent across equatings, the differences themselves are relatively minor when compared to, for instance, the scaled standard error of measurement for SAT-verbal, which is approximately 30 scaled score points.

The pre-equating of Form 3BSA3 was clearly not nearly as successful as the pre-equating for Form 3ASA3. The residual plots show maximum differences in scaled scores resulting from the pre-equating and the operational calibration system or direct link IRT equating of upwards of 20 score points. Once again, the differences between the pre-equating and the intact form linear equating are even greater, particularly in the regions of the formula score scale where the raw to scaled conversion is curvilinear. The differences in scaled score means and standard deviations resulting from the pre-equating and the comparison equatings are much larger than those for Form 3ASA3. The two IRT methods used as criteria produced scaled score summary statistics that are very similar. Unlike the equatings for 3ASA3, scaled score summary statistics produced by the linear equatings are very similar to those produced by the IRT criterion equatings. The scaled score mean resulting from the pre-equating is about ten points greater than the scaled score means resulting from the IRT intact form calibration system, IRT intact form direct link, and intact form linear equatings, which are all within a scaled score point of each other. Once again, the patterns in the

residual plots for the IRT pre-equating and the comparison IRT equatings are the same across both of the single equatings displayed in Figures 8 and 9 (to YSA2 and to 3ASA1) and, hence, the subsequent average equating displayed in Figure 10. The pre-equating results in slightly lower scaled scores at the lower end of the formula score scale and consistently higher scaled scores through the middle and upper end of the formula score scale. The maximum differences occur in all plots around a formula score of 70.

A number of possible explanations were generated for why the 3BSA3 verbal pre-equating results were different from the 3BSA3 comparison results and clearly not of the same quality as the 3ASA3 verbal pre-equating results. Explorations of these possible reasons for the inferiority of the 3BSA3 pre-equating results are reported in a subsequent section of this report.

SAT-mathematical Pre-equating Results

The pre-equatings for 3ASA3 and 3BSA3 mathematical were much like the results for the 3BSA3 verbal pre-equating and, hence, are a matter of concern. The residual plots for 3ASA3 mathematical show maximum differences in scaled scores resulting from the pre-equating and the operational calibration system or direct link IRT equating of upwards of 20 score points. The differences between the pre-equating of 3ASA3 to XSA2 and the intact form linear equating are greater then the differences between the pre-equating and either of the intact form IRT equatings, while the same differences for the 3ASA3 to ZSA1 pre-equating are more comparable and slightly less than the 3ASA3 to YSA2 differences. For both of the single equatings, however, the pre-equating raw to scale conversions are higher

than the intact form linear raw to scale conversions for all possible formula scores. The differences in scaled score means and standard deviations resulting from the pre-equating and the comparison equatings are of the magnitude observed for the pre-equating of 3BSA3 verbal. The two IRT methods used as criteria produced scaled score summary statistics that are very similar and the intact form linear summary statistics are also very close to those for the two intact form IRT methods. The scaled score mean resulting from the pre-equating is about twelve points higher than the scaled score means resulting from the other equatings. Once again, the patterns in the residual plots for the IRT pre-equating and the comparison IRT equatings are the same across both of the single equatings displayed in Figures 11 and 12 and, hence, the subsequent average equating displayed in Figure 13. The pre-equating results in comparable or slightly lower scaled scores at the very lower end of the raw score scale (i.e., raw scores less than zero) and consistently higher scaled scores throughout the remainder of the raw score scale. This is not unlike the pattern observed for 3BSA3 verbal, except that in that pre-equating the point on the formula score scale where the pre-equating began to consistently produce higher scaled scores was slightly higher, around 10 or 15. For this pre-equating, the maximum differences in scaled scores occur in all IRT residual plots around a formula score of 50 to 55.

The pre-equating results for 3BSA3 mathematical are almost a carbon copy of the results for 3BSA3 verbal and, thus, are also very similar to the results for 3ASA3 mathematical. The residual plots show maximum differences in scaled scores resulting from the pre-equating and the operational calibration system or direct link IRT equatings of upwards of 25 score

points. The differences between the pre-equating and the intact form linear equating are even greater, particularly in the regions of the formula score scale where the raw to scaled conversion is curvilinear. The differences in scaled score means and standard deviations resulting from the pre-equating and the comparison equatings are in the same direction as those for 3ASA3 mathematical and 3BSA3 verbal, but are of a slightly smaller magnitude. The two IRT methods used as criteria produced scaled score summary statistics that are very similar, but this time the intact form linear summary statistics, in particular, the scaled score mean, was somewhat discrepant from the two IRT methods. The scaled score mean resulting from the pre-equating is about eight points higher than the scaled score mean resulting from the other IRT equatings. Once again, the patterns in the residual plots for the IRT pre-equating and the comparison equatings are the same across both of the single equatings displayed in Figures 14 and 15 and, hence, the subsequent average equating displayed in Figure 16. The pre-equating results in lower scaled scores at the lower end of the raw score scale (i.e., raw scores less than 10) and consistently higher scaled scores throughout the remainder of the raw score scale. In this pre-equating, the maximum difference in scaled scores occurs in all IRT residual plots around a formula score of 45 to 50. It also should be noted that the patterns in the pre-equating residual plots for 3BSA3 mathematical are very similar to those displayed for 3BSA3 verbal and 3ASA3 methematical. For all three pre-equatings, the raw to scaled conversions resulting from the pre-equatings are consistently higher than the conversions resulting from the comparison equatings for formula scores of ten or greater.

Supplemental SAT-verbal Investigations and Results

In this section, the results of investigations into possible reasons for the problematic 3BSA3 verbal pre-equating are reported. A number of possible explanations were generated, and then each was investigated individually.

One possible explanation to be considered for the 3BSA3 verbal pre-equating results has to do with practice effects generated from the manner in which the test sections were sequenced. In other words, for 3ASA3 verbal, there may have been more or less of a balancing effect of the sequencing of the operational final form section that the pretest section was built to parallel and the pretest section itself (perhaps in about 50% of the final form - pretest combinations represented in Figure 1 the operational section appeared first and in the other 50% of the combinations the pretest section appeared first), while for 3BSA3 the balancing may not have occurred. However, upon closer consideration, if pretest practice effects had indeed taken place for Form 3BSA3, the pre-equating results would have been in exactly the opposite direction from the comparison results then what actually occurred. In other words, if the pretest section followed the operational section in a disproportionate number of cases, and practice effects occurred, then the items as they appeared in pretest form would have appeared easier than in final form. The pre-equating results would have been consistently lower in the upper part of the formula score scale than the comparison equating results. Just the opposite took place, and hence, it must be the case that practice effects can be ruled out as an explanation for why the 3BSA3 pre-equating results were so different from the comparison equating results and also from the 3ASA3 pre-equating

results. Out of interest, an investigation of the sequencing of the verbal sections of 3ASA3 and 3BSA3 was performed anyway. For 3BSA3, in 65% of the final form - pretest combinations, the pretest followed the operational section; for 3ASA3, this was true 64% of the time. Thus, while there appeared to be at least the potential for practice effects to occur for both forms, the pre-equating results clearly rule them out.

Two other potential explanations for differences in the verbal pre-equating results have to do with equating samples and LOGIST calibration runs. These are:

1. The use of two different equating samples with the 3BSA3 intact form calibration system and direct link equatings. In doing the 3ASA3 intact form calibration system and direct link equatings, the same equating section, fm, was in common with old forms XSA2 and YSA3, and hence the same sample, and subsequent set of parameter estimates, could be used for both equatings. This was not true for 3BSA3 in that fk was in common with YSA2 and fw with 3ASA1. This necessitated the use of two different samples, and hence, two different sets of item parameter estimates (both sets taken from the SAT IRT Scale Drift Study) to perform the equatings.
2. The use of different versions of the LOGIST program to generate item parameter estimates. For 3ASA3, both the pretest and the final form parameter estimates were generated from the current version of LOGIST, and this is also true of the 3BSA3 pretest parameter estimates. To save on calibration costs, the 3BSA3 final intact form parameter estimates were recovered from the SAT IRT Scale Drift Study (Petersen, et al, 1983) run on a different version of LOGIST.

It is possible that the updating and refinement of the LOGIST program caused subtle differences in parameter estimates, which collectively caused the differences seen in the residual plots for 3BSA3.

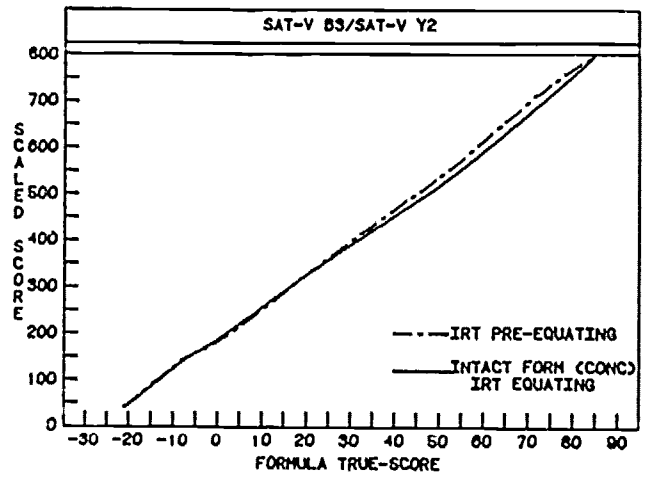
The possible explanations above implicitly assume that it was not the pre-equating for 3BSA3 that was somehow faulty, but instead the comparison IRT equatings. To investigate whether or not it is reasonable to explain the differences in pre-equating results this way, the following was done. The operational final form-equating section combinations needed to equate intact final form 3BSA3 to old forms YSA2 and 3ASA1 (see Figure 3) were run together in one large LOGIST run, using the current version of LOGIST, and the intact final form equating redone. As a result, the parameter estimates for the 3BSA3 pre-equating and the 3BSA3 intact final form equating were generated using the same version of LOGIST. Further, by running the data for 3BSA3 and the two old forms concurrently, there was no need for scaling parameter estimates (all parameter estimates needed in the equating are automatically on the same scale) and only one set of 3BSA3 final form parameter estimates were used in the equating (unlike the previous IRT comparison equatings). In sum, the results of equating intact final form 3BSA3 to the old forms using the parameter estimates from the concurrent LOGIST run should provide the best criterion possible for evaluating the 3BSA3 pre-equating results.

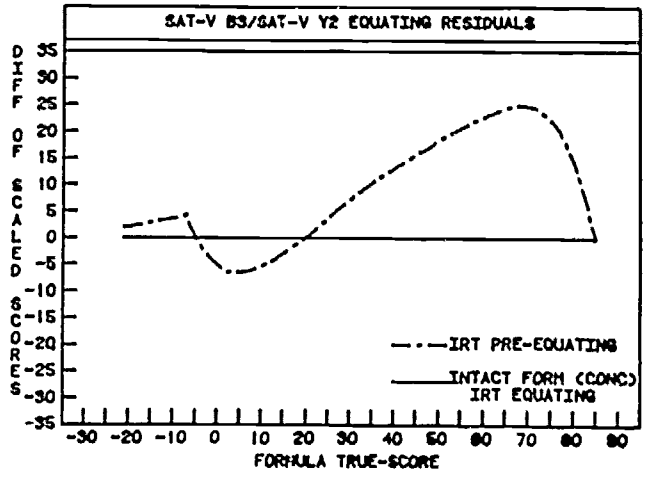
A comparison of the 3BSA3 pre-equating results to this new IRT comparison equating is presented in Figure 17 for each of the single equatings and the average equating. The new comparison equating has been labeled intact form concurrent equating in this figure. Information on the

Equating Plot

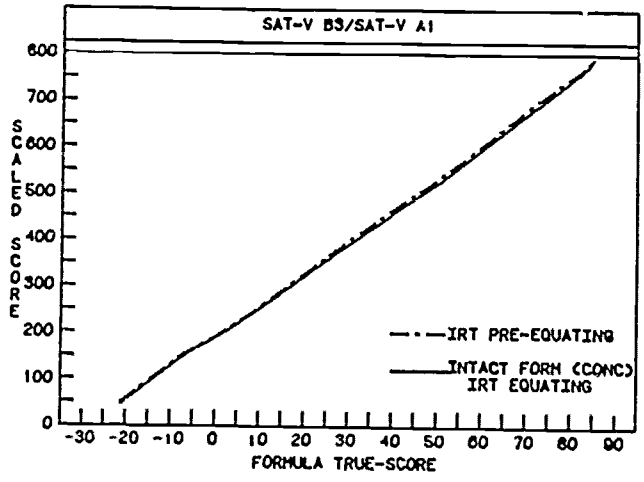
Residual Plot

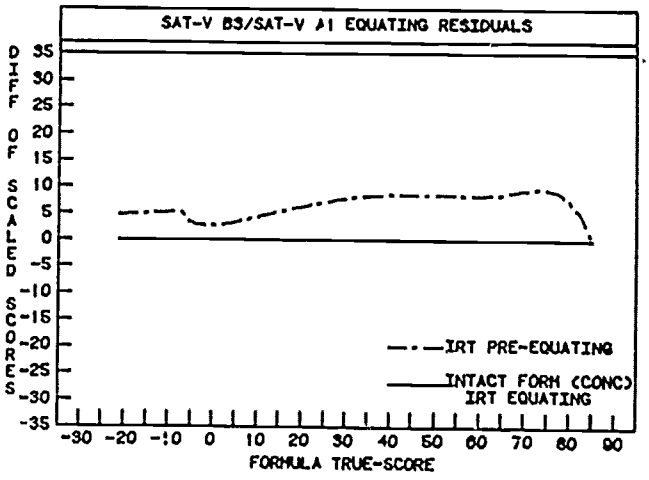
YSA2



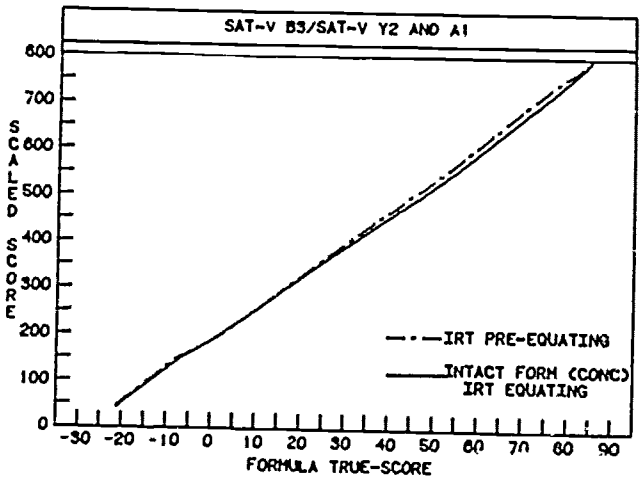


3ASA1





YSA2 and 3AS1



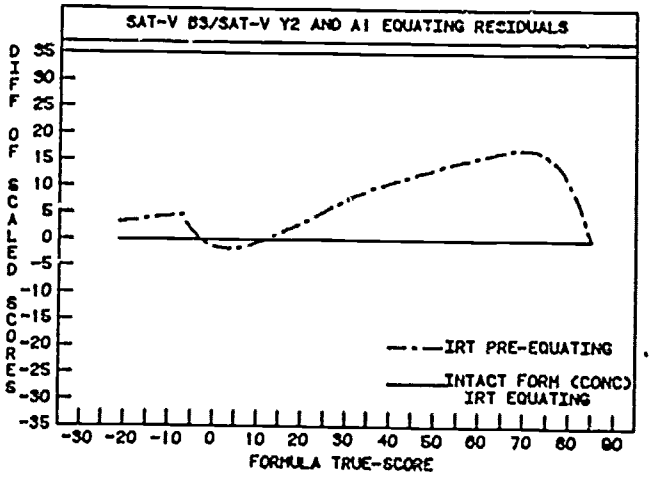


Figure 17: SAT-verbal Form 3BSA3 equated to SAT-verbal Form YSA2, Form 3ASA1, and Forms YSA2 and 3ASA1 - Plots of 1) IRT pre-equating raw to scaled score transformation compared to intact final form concurrent IRT equating raw to scaled score transformation, and 2) differences between scaled scores (IRT pre-equating - intact final form concurrent IRT equating) resulting from the equatings.

scaled score summary statistics resulting from this new equating and the other previously described is presented in Table 8.

The results presented in Figure 17 clearly lead to the conclusion that it is not the comparison IRT equatings for Form 3BSA3 that are faulty, but rather the Form 3BSA3 pre-equating results. The data presented in Figure 17 show differences between the IRT pre-equating and the intact form concurrent IRT equating that are comparable to the differences in the residual plots using the other intact form comparison equatings. Thus the possible explanations for differences in equating results based on the use of different versions of LOGIST and multiple sets of parameter estimates, generated from the IRT Scale Drift Study (Petersen, et al, 1983) must be discounted.

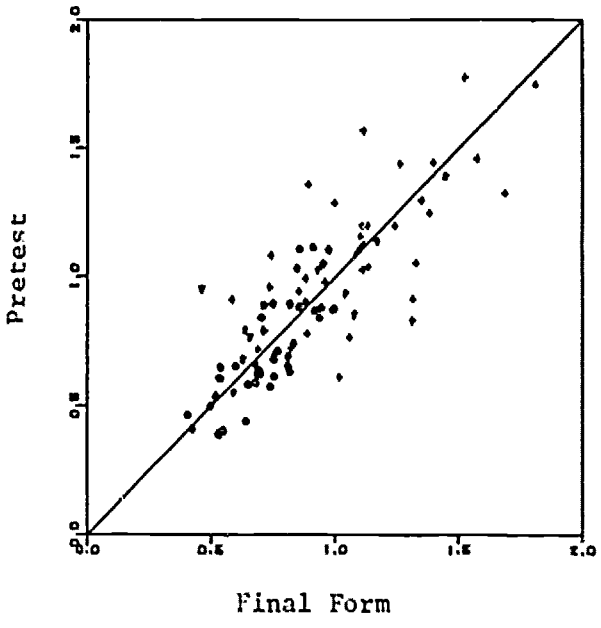
The only other possible explanation for the differences between the Form 3BSA3 verbal pre-equating and intact form comparison equating results has to do with the quality of the parameter estimates for the 85 3BSA3 items when they appeared in pretest form. In order for the equatings to be as discrepant as they are, the pretest and final form parameter estimates for certain of the items must be quite different. The following methods were used to compare these two sets of parameter estimates in an attempt to both observe general differences and trends in the individual parameter estimates and locate those items for which the pretest parameter estimates were problematic. First, two-way plots of pretest and final form item discrimination (a), lower asymptote (c), and item difficulty (b) parameter estimates were prepared. Figure 18 contains the three plots relevant to 3BSA3 verbal. Second, a mean absolute difference between the item response functions for each item, where the functions were generated using the

Table 8

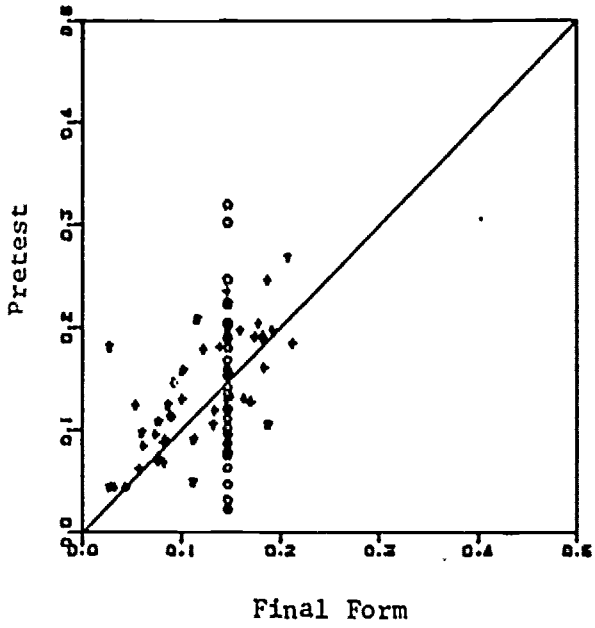
Scaled Score Summary Statistics Resulting from Application of Five Equating Methods
SAT-verbal Sections of Form 3BSA3

Form	N		IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Intact Form (Concurrent Run)	IRT Pre-equating
3BSA3	253,354	M	430.25	430.42	431.42	431.54	440.39
		S.D.	105.99	105.57	106.53	105.86	110.55

Item discrimination (a)
parameter estimates



Lower asymptote (c)[1]
parameter estimates



Item difficulty (b)
parameter estimates

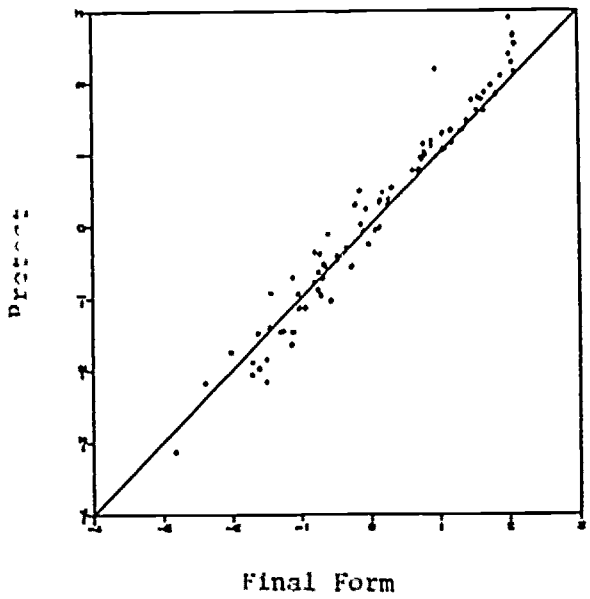


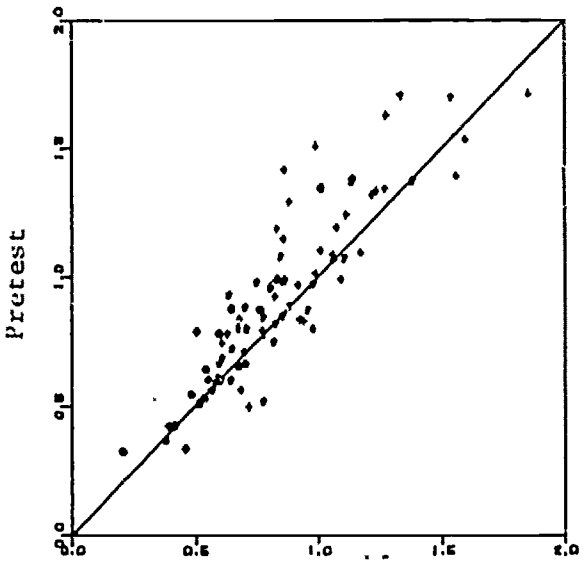
Figure 18: Two-way plots of pretest and final form parameter estimates for the 85 3BSA3 verbal items.

[1] Items lying along the vertical line with an abcissa of approximately .15 are items whose final form lower asymptotes were fixed at a common value.

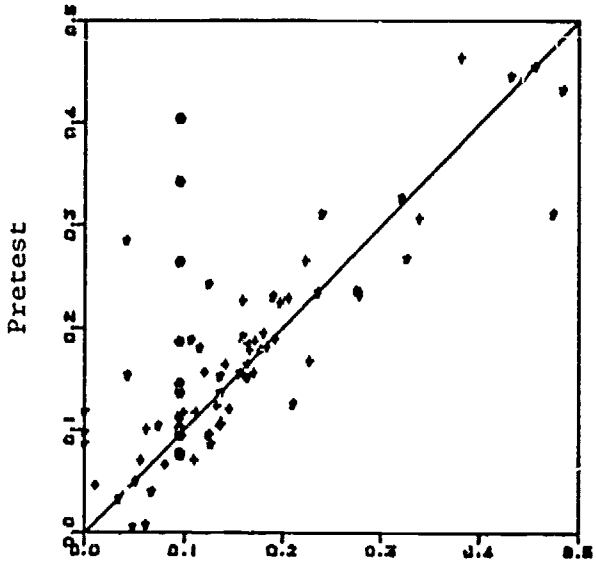
pretest and the final form parameter estimates, were obtained. Using all individuals in the sample taking Form 3BSA3 when calibrated as an intact final form, the absolute difference between the item response functions for each person (i.e., value of θ) was obtained and then averaged across people. Items having the largest mean absolute difference values were then located. The above analyses were also done for the two sets of Form 3ASA3 item parameter estimates so that the discrepancies between parameter estimates for 3ASA3, where the pre-equating results were more acceptable, could be compared to the 3BSA3 discrepancies. Figure 19 contains the two- way plots of individual parameter estimates for Form 3ASA3.

Looking at the two-way plots in Figure 18, one important result becomes evident. In the plot of the pretest and final form item difficulty parameter estimates, there are a much larger number of individual points lying above the diagonal than below. Points lying on the diagonal are items that have no difference between pretest and final form difficulty parameter estimates. Points above the diagonal indicate items that were estimated to be more difficult in the pretest than in the final form. Of the 85 3BSA3 items, 59 (69%) were estimated to be more difficult in the pretest than in the final form. For 3ASA3 on the other hand (see Figure 19), there is a better balance of items lying above and below the diagonal of the two-way plot of item difficulty estimates. For that form, 45 of the 85 items (53%) were estimated to be more difficult in the pretest than in the final form. Two-way plots of item discrimination and lower asymptote parameter estimates in Figures 18 and 19, while indicating a good deal more variability in individual item parameter estimates than the two-way item difficulty plots, also demonstrate the expected balance of points above and below each

Item discrimination (a)
parameter estimates



Lower asymptote (c)
parameter estimates



Item difficulty (b)
parameter estimates

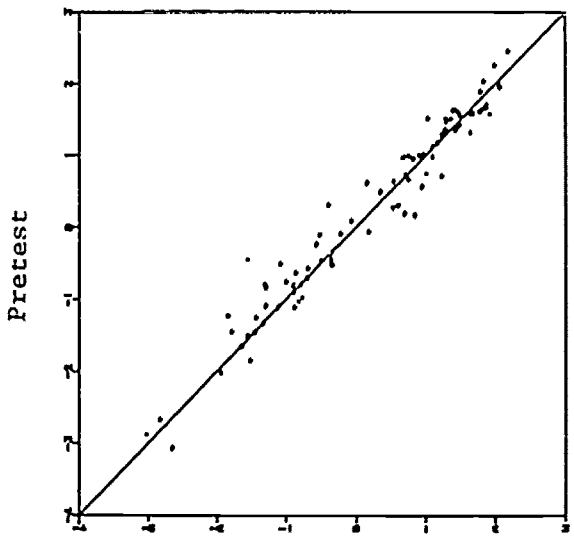


Figure 19: Two-way plots of pretest and final form parameter estimates for the 85 3ASA3 verbal items.

diagonal. The one exception is the two-way plot of 3ASA3 pretest-final form item discrimination estimates. A predominant number (60, or 70%) of the items were estimated to be more discriminating in the pretest than in the final form.

Using the mean absolute difference between the item response functions as a criterion for selection of problematic items, thirteen items from 3BSA3 and twelve from 3ASA3 were identified. Upon inspection of these two subsets of problem items, they were found to differ considerably in characteristics. Of the thirteen items identified for Form 3BSA3, eleven were reading comprehension items. Of the eleven, four were based on the same passage and three on another passage. The remaining four reading comprehension items were single items based on four different passages. Of the twelve items identified for Form 3ASA3, four were reading comprehension items (two from one passage, the other two from two other passages different from the first), four were antonym items, three were analogies, and one was a sentence completion item. Of the thirteen 3BSA3 items identified, twelve were more difficult when given in a pretest than in the final form. Of the twelve 3ASA3 items, seven were more difficult when given in a pretest and five when given as part of the intact final form.

Upon closer inspection of the eleven reading comprehension items from 3BSA3 exhibiting large absolute differences in item response functions, it was found that nine of these items came from pretests in which the passage they were linked to was located in the final position of the pretest section. For all but one of these items, the item as it appeared in the pretest was more difficult, sometimes considerably more, than when it appeared in the final form. For the lone exception, a word was deleted from

the correct response (the only such occurrence on either 3ASA3 or 3BSA3) between the time when the item was given in a pretest and the time it appeared in the final form. This word also appeared as a key word in the text of the passage and it may be hypothesized that it acted as a clue to the correct response, thus explaining the increase in difficulty upon removal of the word from the correct response when the item appeared in the final form. Figure 20 contains plots of the item response functions based on pretest and intact final form parameter estimates for the thirteen problematic 3BSA3 items, identified by item type. For the reading comprehension items (numbered 1-11), items numbered 1-4 are all based on the same reading passage, as mentioned earlier, and items numbered 5-7 are based on the other passage discussed. Reading comprehension item number 11 is the item in which the word was deleted from the correct response between when the item was given in pretest and in final form. The remaining two problematic items are presented in Figure 20 after the reading comprehension items; one of the items is an analogy item and the other is an antonym item.

Upon inspection of the four problematic reading comprehension items identified in Form 3ASA3, it was determined that, exactly like the situation for Form 3BSA3, the items were located in pretests where the passage they were linked to was located last in the pretest section. All four items were also more difficult when they appeared in the pretest section than in the final form.

On the basis of the above data, it may be hypothesized that either something approaching a fatigue factor, such as that found in the Kingston and Dorans (1982) study, is being exhibited in the responses of candidates to reading comprehension items based on passages located at the end of

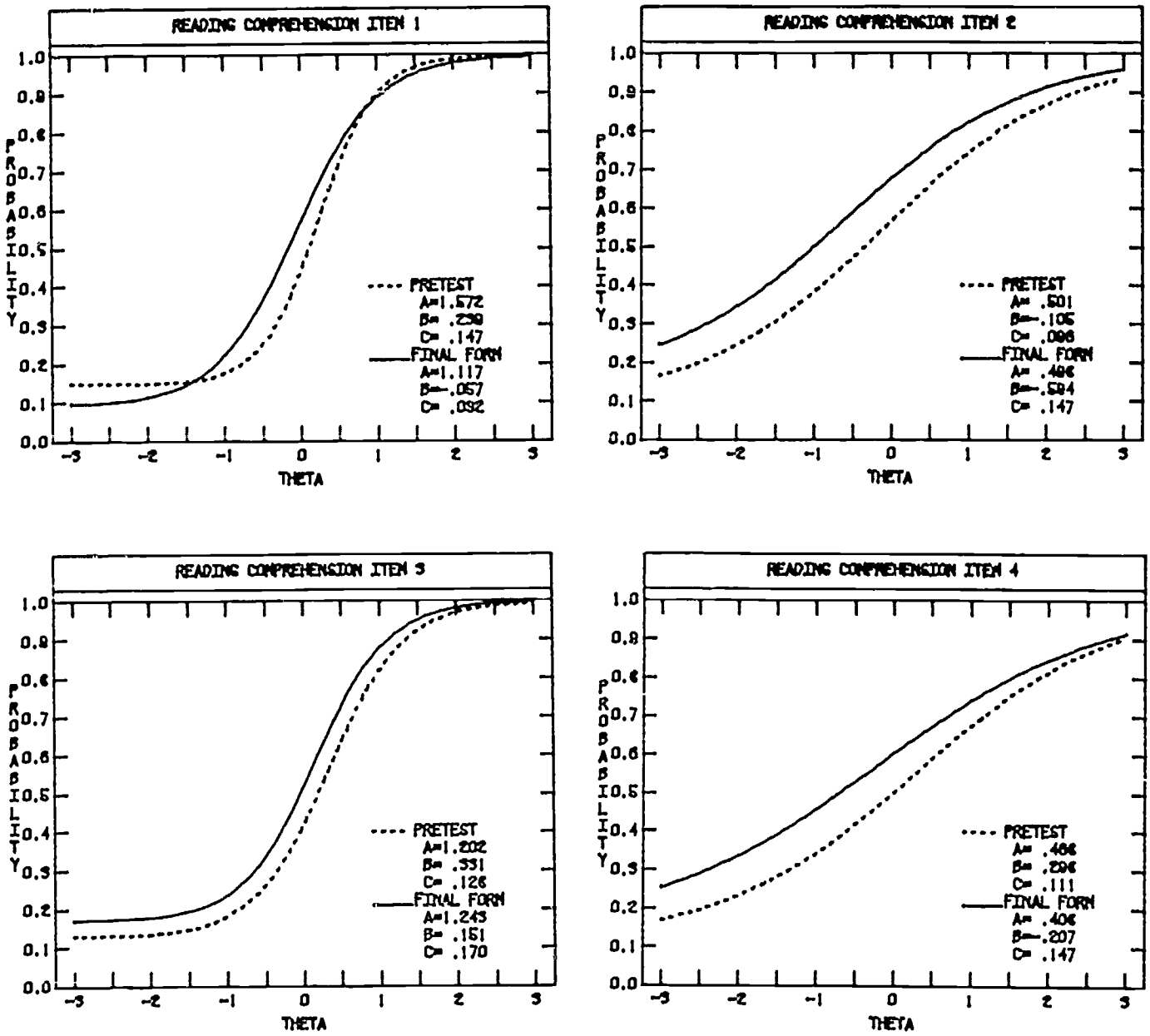


Figure 20: Plots of item response functions based on pretest and intact final form item parameter estimates for thirteen problematic Form 3BSA3 items.

BEST COPY

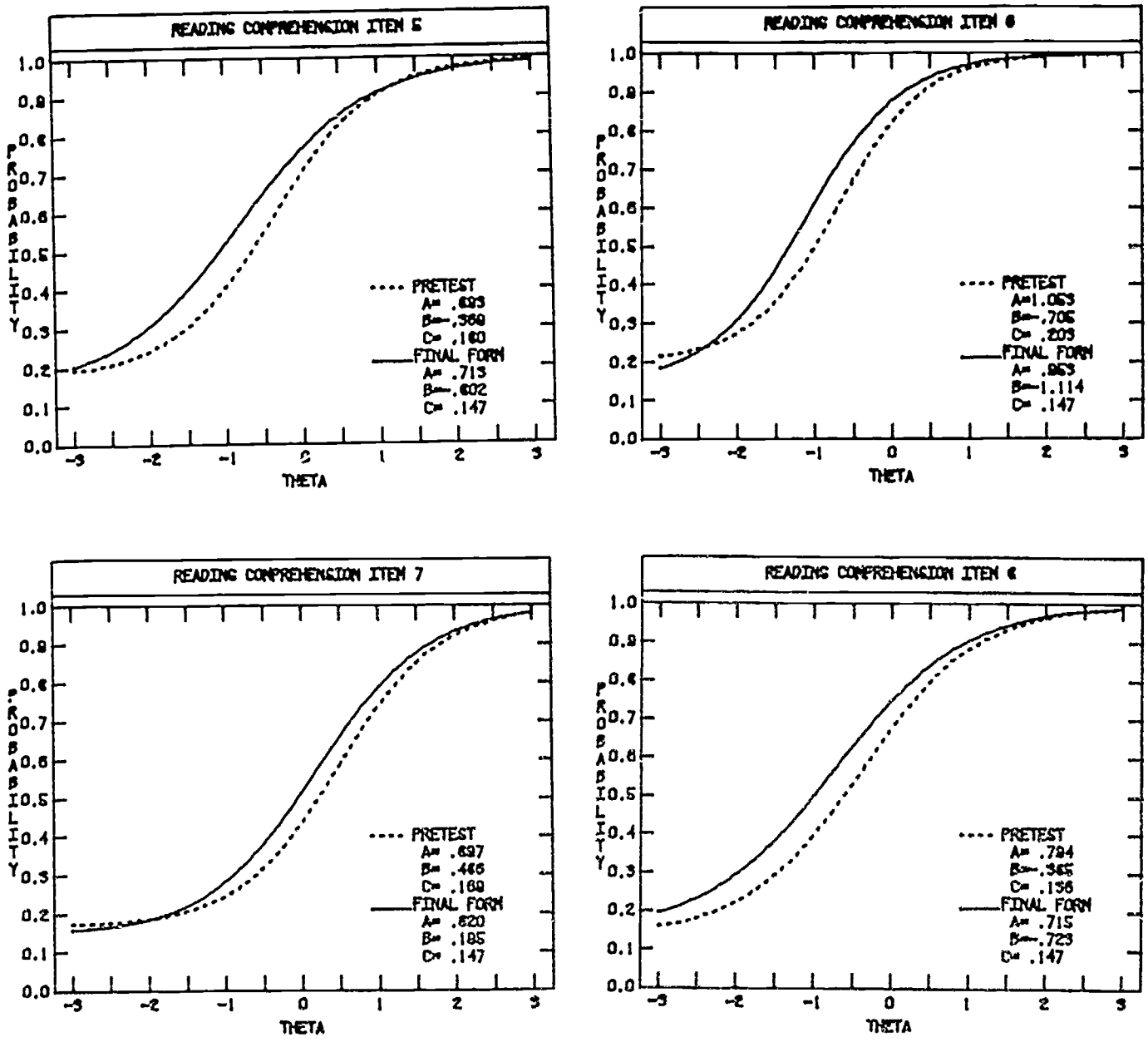
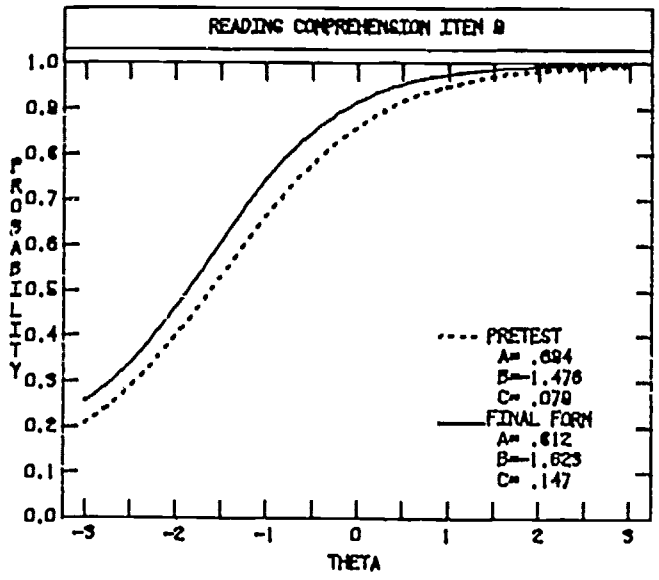
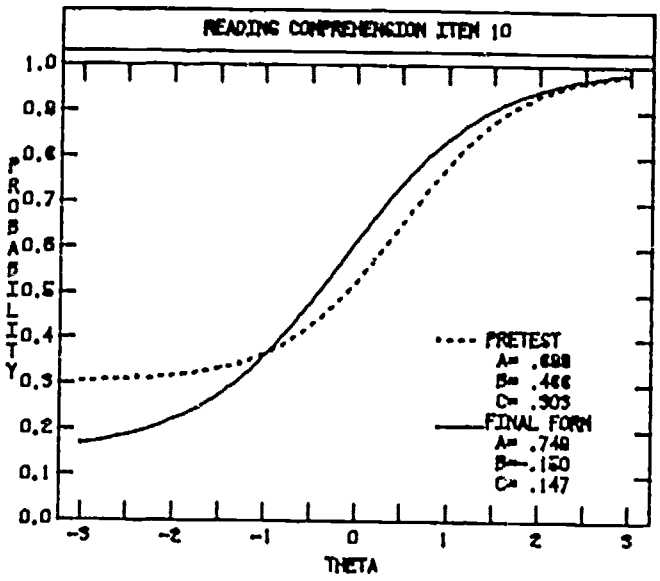
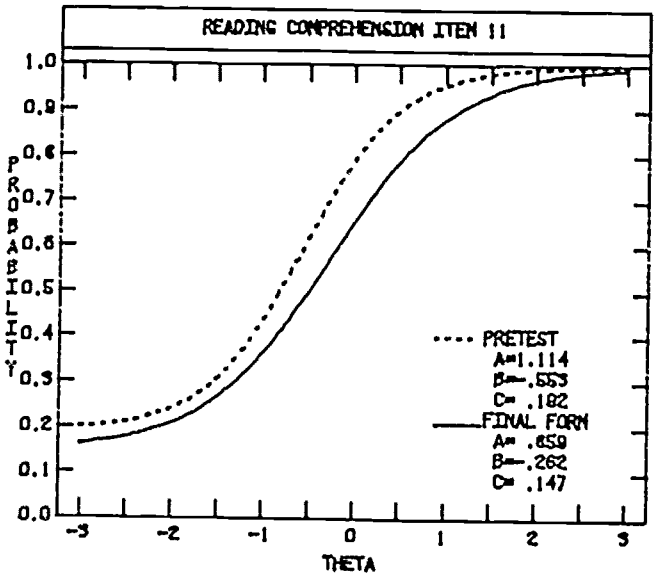


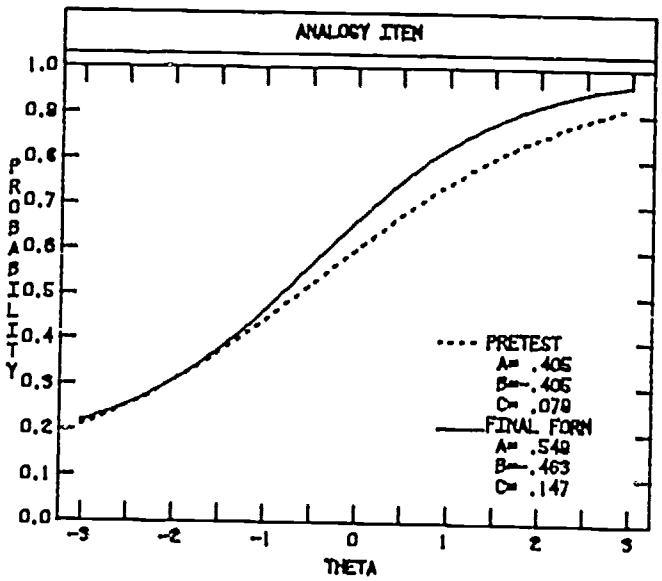
Figure 20: Plots of item response functions based on pretest and intact final form item parameter estimates for thirteen problematic Form 3BSA3 items.

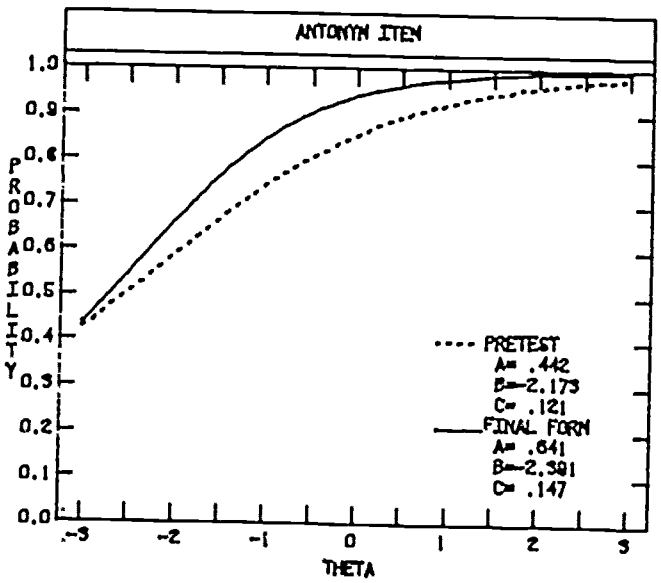
BEST COPY











BEST COPY

Figure 20: Plots of item response functions based on pretest and intact final form item parameter estimates for thirteen problematic Form 3BSA3 items.

pretest sections or, because of lack of time, random responses are being supplied by these candidates to certain of the questions based on this last passage. In either case, the items are more difficult in the pretest than they are in the final form, where due to passage location, a fatigue factor or the supplying of random responses is not occurring. The data on the reading comprehension items from both 3BSA3 and 3ASA3 are consistent with this statement.

If the above is happening to pretest reading comprehension items based on passages located at the end of pretest sections, one might be concerned about whether there are large discrepancies in parameter estimates between pretest and final form for reading comprehension items in the intact final form based on passages at the end of the SAT-verbal 45 item and 40 item sections. The 45 item SAT-verbal sections do not end with reading comprehension items, but the 40 item sections do. For 3ASA3, the passage upon which the last set of reading comprehension items (items 36-40) were based was also located in the final position in the pretest. Two of the five items still demonstrated discrepancies large enough to be included in the overall set of twelve items discussed earlier. For 3BSA3, the passage upon which the last set of reading comprehension items (items 36-40) were based was not located at the end of the pretest section. It was, however, the only reading comprehension passage in the pretest, and one of these last five items (36-40) in 3BSA3 did exhibit large discrepancies in pretest-final form parameter estimates. Thus it would appear that while the outcome in terms of parameter estimate discrepancies for reading comprehension items located at the end of SAT-verbal sections is not as clear cut as for comparable pretest reading comprehension items, there is still cause for concern.

The effect on equating of having, in particular, reading comprehension pretest item difficulties estimated to be higher than they are when estimated on intact final form data is predictable, and demonstrated in the Form 3BSA3 pre-equating results. If the same items are more difficult in the first "test" (made up of pretest items) than in the second (made up of the items in the intact final form), then the same raw score on both "tests" should result in a higher scaled scores on the first "test" than the second. This appears to be exactly what is happening with the 3BSA3 pre-equating results. The reading comprehension items that have been specifically discussed are contributing to the 3BSA3 pre-equating results along with the other items, not specifically discussed, but lying above the diagonal of the two-way plot of item difficulty estimates in Figure 18. It is these other items for which there is another level of concern. The discrepancies between the item difficulties for these items are often only slight, but indicating in each case that the item was estimated as being more difficult in the pretest. (The mean of the pretest item diffculties for all items was .198 while the mean of the final form item difficulties was .107.) Collectively, these items will exert an influence, in conjunction with the more discrepant items specifically described above, on the 3BSA3 pre-equating results. For 3ASA3, on the other hand, there is more of a balancing effect of the discrepancies between pretest and intact final form parameter estimates and the result is that the pre-equating and the intact final form comparison IRT equatings more closely coincide.

Supplemental SAT-mathematical Investigations and Results

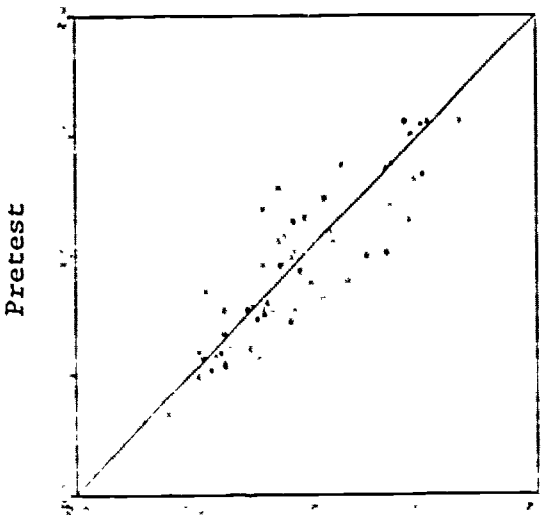
Because the verbal and mathematical pre-equatings were performed sequentially, and the verbal portion was completed before the mathematical portion was begun, a number of the supplemental investigations done for the verbal pre-equatings were not of concern when the mathematical results were obtained. In the case of certain of the potential problems forcing the supplemental investigations for the verbal pre-equatings, the problems were corrected before the mathematical portion of the study was begun. In the case of other supplemental verbal investigations, such as the investigation of practice effects, the results of these investigations were such as to rule out the problems investigated as explanations for the poor mathematical pre-equatings.

After consideration of the verbal supplemental investigations, it appeared for the mathematical pre-equatings, as was also the case for verbal, that the one possible explanation for the difference between the mathematical pre-equatings and the intact form equating results had to do with the quality of the parameter estimates for the 3ASA3 and 3BSA3 items when they appeared in pretest form. In order for the equatings to be as discrepant as they were, the pretest and final form parameter estimates for certain of the items must be quite different. Further, based on the supplemental investigations for 3BSA3 verbal and the scaled score summary statistics for 3ASA3 and 3BSA3 mathematical, the direction of these differences in parameter estimates became evident. Either certain of the items were estimated to be much more difficult in pretest then in final form or there was an overwhelming trend, with individual differences perhaps slight, for the items to be estimated as being more difficult when given in pretest form.

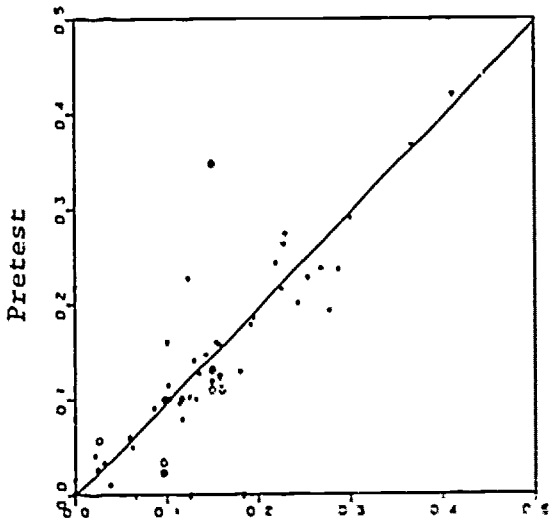
The methods used to investigate the verbal pretest and final form parameter estimates were also used here. Two-way plots of pretest and final form item discrimination, lower asymptote, and item difficulty parameter estimates were prepared. A mean absolute difference between item response functions for each item, based on pretest and final form parameter estimates, was obtained and items having the largest differences were located. These procedures were described in greater detail for the SAT-verbal investigations on page 52 of this report.

Figure 21 contains the two-way plots for 3ASA3 mathematical and Figure 22 contains comparable data for Form 3BSA3. As was the case for the two-way plots for 3BSA3 verbal, one important result becomes evident upon looking at these plots. In the plots of the 3ASA3 and 3BSA3 pretest and final form item difficulty parameter estimates, there are a much larger number of individual points lying above the diagonal than below. Points above the diagonal indicate items that were estimated to be more difficult in pretest than in final form. Of the 60 3ASA3 mathematical items, 45 (75%) were estimated to be more difficult in pretest than in final form; for the 60 3BSA3 mathematical items, 42 (70%) were estimated to be more difficult. For 3ASA3, the mean of the pretest item difficulties was .355, while the mean of the final form item difficulties was only .227. Comparable figures for 3BSA3 are .365 and .253. The two-way plots of item discrimination and lower asymptote parameter estimates in Figures 21 and 22, while again indicating a good deal more variability in individual parameter estimates than the item difficulty parameter estimates, demonstrate the more expected balance of points above and below each diagonal. For instance, for the 3ASA3 two-way item discrimination plot, 27 (45%) of the points lie above the

Item discrimination (a)
parameter estimates



Lower asymptote (c)
parameter estimates



Item difficulty (b)
parameter estimates

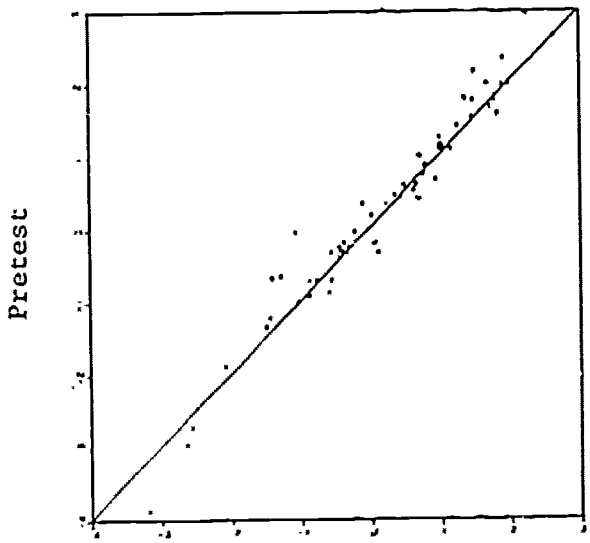
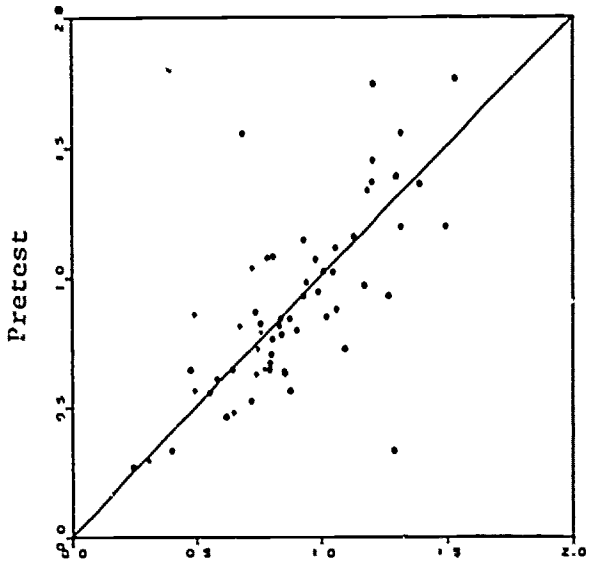
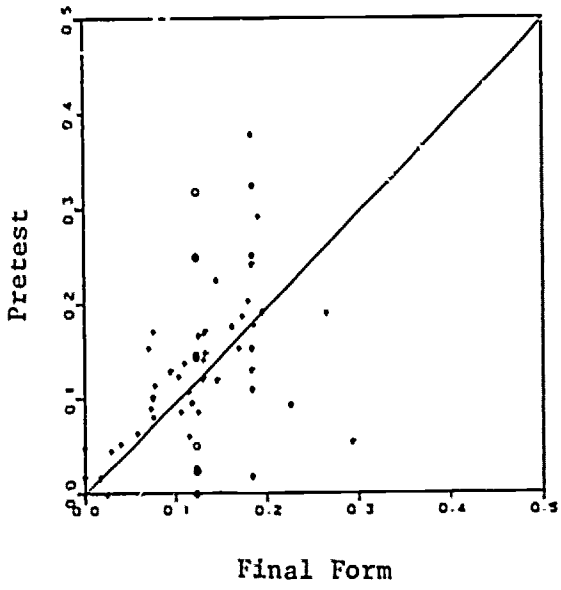


Figure 21: Two-way plots of pretest and final form parameter estimates for the 60 3ASA3 mathematics items.

Item discrimination (a)
parameter estimates



Lower asymptote (c)
parameter estimates



Item difficulty (b)
parameter estimates

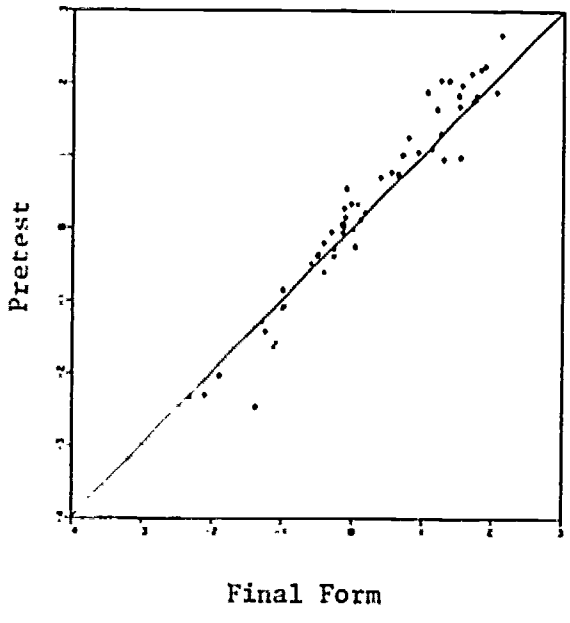


Figure 22: Two-way plots of pretest and final form parameter estimates for the 60 3BSA3 mathematics items.

diagonal and 33 (55%) below. For the 3BSA3 two-way item discrimination plot, 30 (50%) of the points lie above the diagonal and 30 (50%) below.

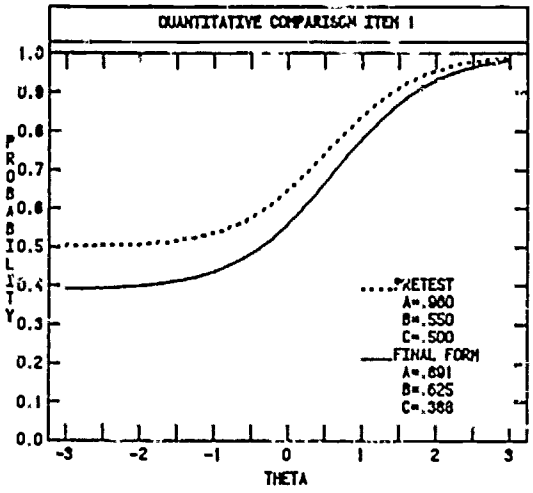
Using the mean absolute difference between the item response functions as a criterion for selection of problematic items, ten items from 3ASA3 and fourteen items from 3BSA3 were located that had large differences, of the magnitude used to identify the verbal problematic items. The plots of the item response functions based on pretest and intact final form parameter estimates for these items are contained in Figures 23 and 24. The items are arranged sequentially in these plots in descending order by mean absolute difference in item response functions (i.e., quantitative comparison item 1 in Figure 23 had the largest absolute difference of all 3ASA3 items and regular math item 10 had the smallest absolute difference for all items deemed problematic). Such a rank ordering was not done for the 3BSA3 verbal problematic items because an explanation for many of the problems was readily apparent from the item type and the positioning of the items, unlike the situation here.

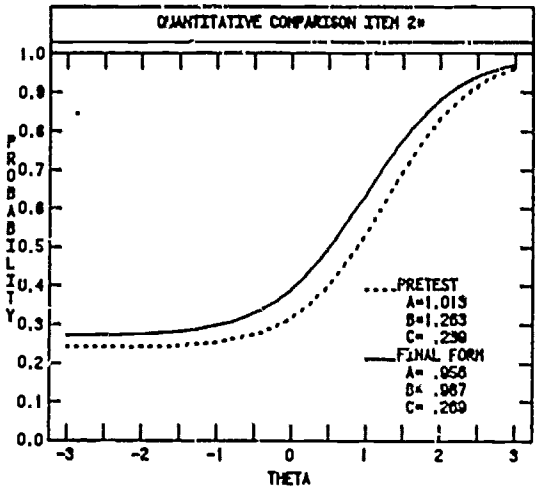
Of the ten 3ASA3 mathematical items having large absolute differences in item response functions, six were four-choice quantitative comparison items (30% of the total 20 items) and four were five-choice regular math items (10% of the total 40 items). Nine of the ten items had item difficulty estimates that were larger when given in a pretest than in the intact final form. The one exception was quantitative comparison item one, which was easier when given in the pretest than in the final form. Of the fourteen 3BSA3 mathematical items having large absolute differences in item response functions, five were quantitative comparison items (25% of the total 20) and nine were regular math items (22.5% of total 40). Thirteen of the fourteen

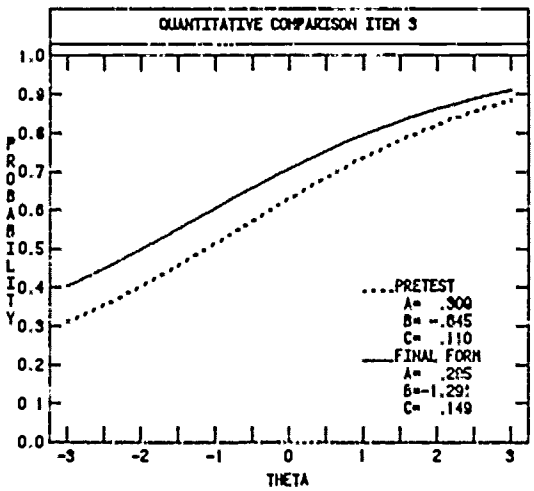
items were estimated to be more difficult in pretest than in final form. For both 3ASA3 and 3BSA3, the items having the largest mean absolute differences came from a number of different pretests contained in a number of different LOGIST runs depicted in Figure 2. In other words, poorly estimated parameters from one or a small number of mathematical pretests are not what is responsible for the discrepancies located.

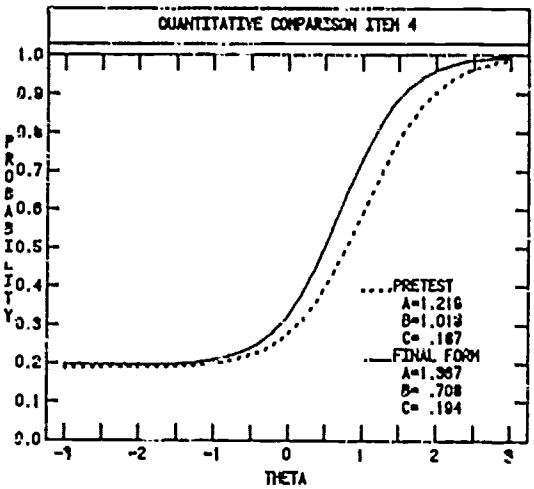
Because the 3BSA3 verbal pre-equating results could be, in large part, explained by the positioning of reading comprehension items at the end of verbal pretests, the twenty four items from 3ASA3 and 3BSA3 mathematical that were deemed problematic were carefully studied to see if position effects could be used as an explanation for the differences observed in item response functions. Of the ten 3ASA3 items, the positioning of the item in the pretest could be used as an explanation for the differences in item response functions, or item difficulty estimates, for two of them. Of the fourteen 3BSA3 items, position effects can be used as an explanation for eight of them. Table 9 contains information on the positioning of these items in the pretest and final form. They are also identified by an asterisk in Figures 23 and 24.

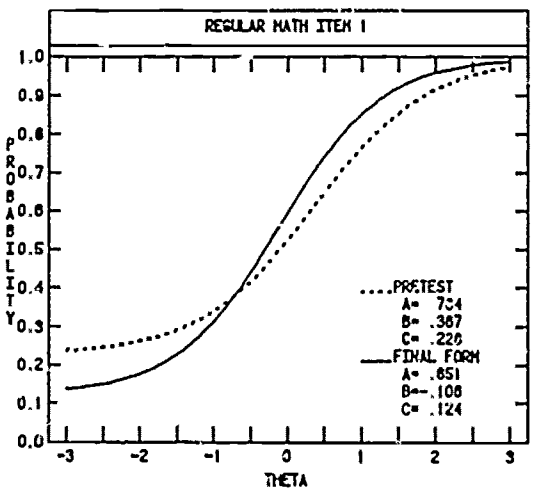
Unlike the case for 3BSA3 verbal (and also, to a certain extent, 3ASA3 verbal), the effects of the positioning of an item in the pretest or in the final form section could not be used to explain a majority of the differences seen in item response functions or item difficulty estimates. Of the 24 total items exhibiting large differences in item response functions, only ten of them could have their differences explained in any way by the positioning of the item in the pretest. Six of the items for which position effect explanations could be posited were regular math items







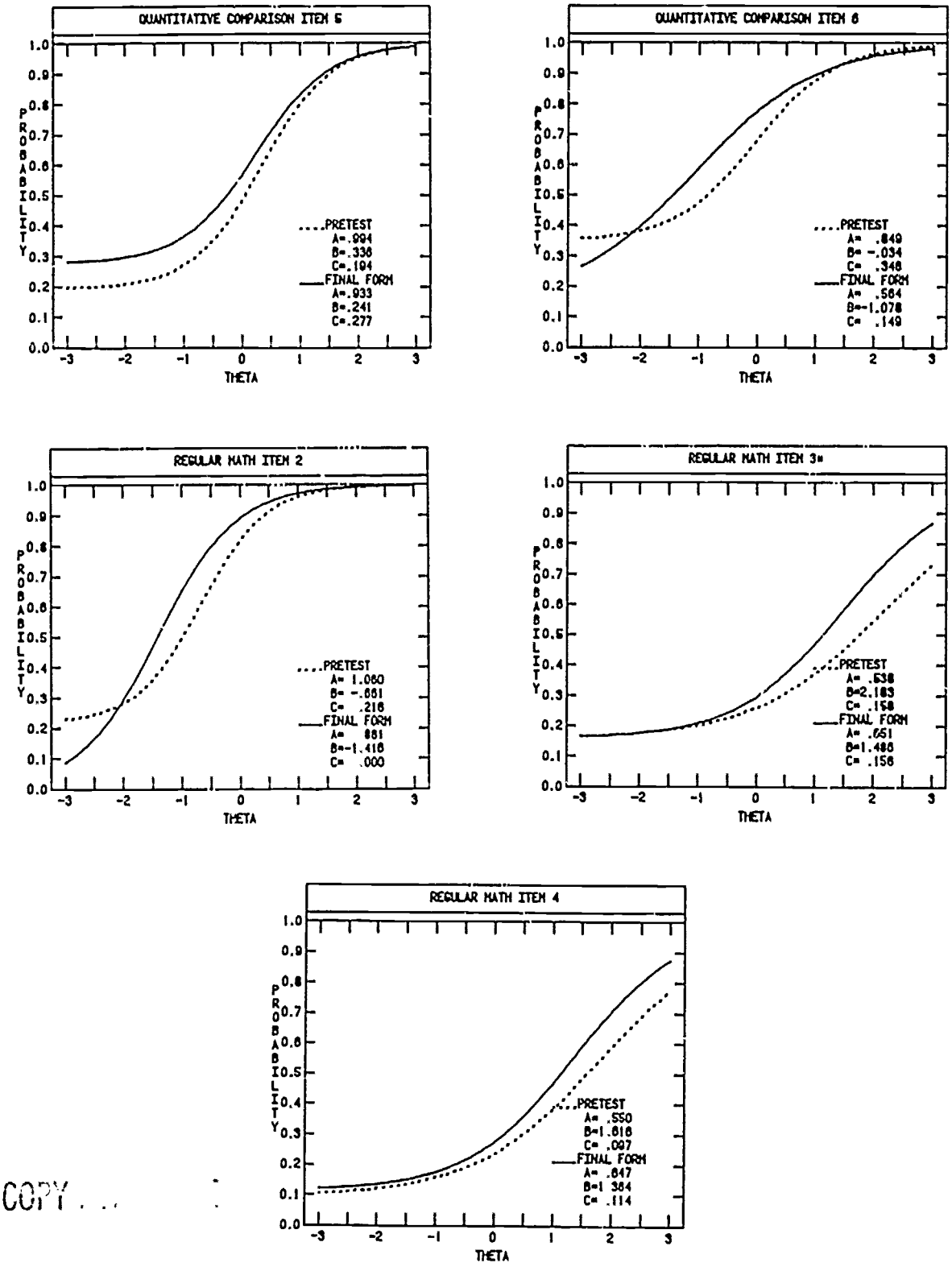




BEST COPY

Figure 23: Plots of item response functions based on pretest and final form item parameter estimates for ten problematic Form 3ASA3 mathematical items.

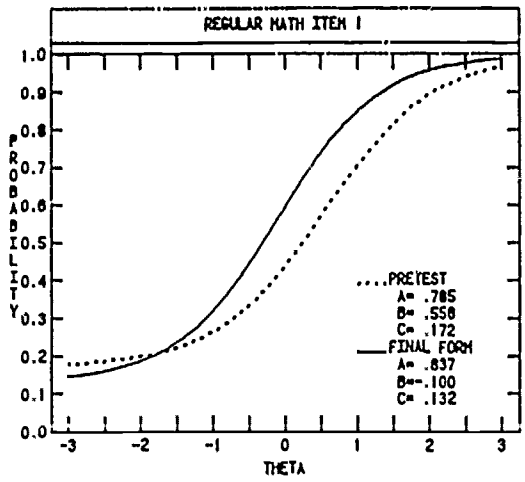
*Item for which difference in pretest and final form item response functions can potentially be explained by item position effects.

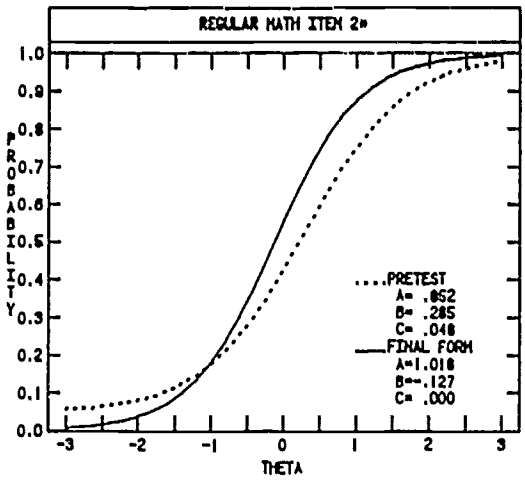


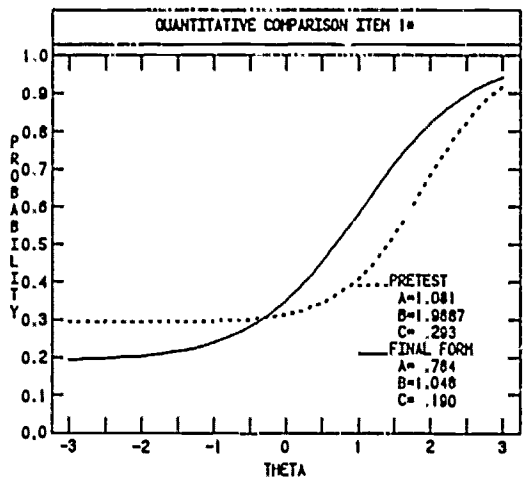
BEST COPY . . .

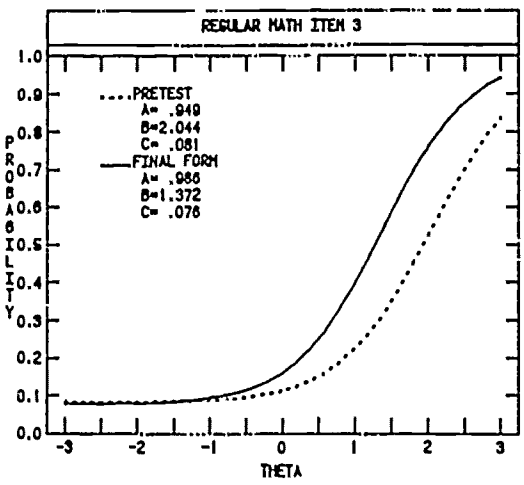
Figure 23: Plots of item response functions based on pretest and final form item parameter estimates for ten problematic Form 3ASA3 mathematical items.

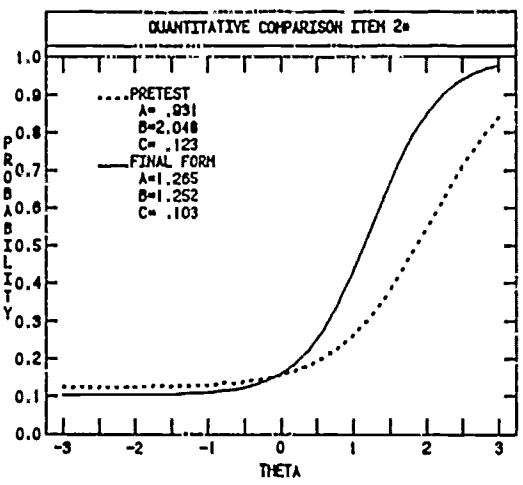
*Item for which difference in pretest and final form item response functions can potentially be explained by item position effects.







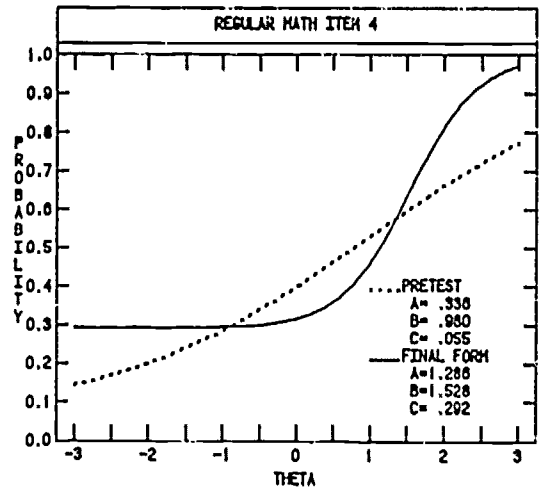


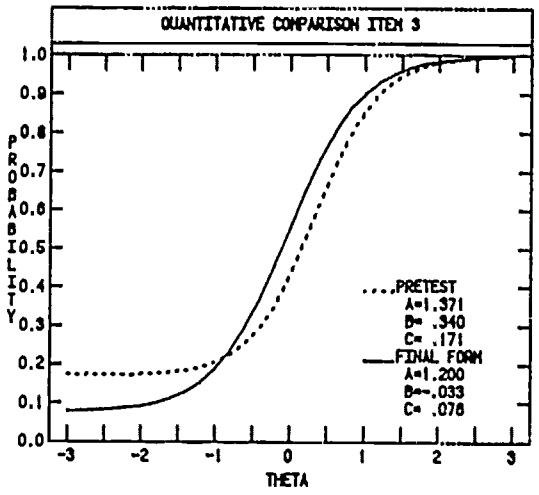


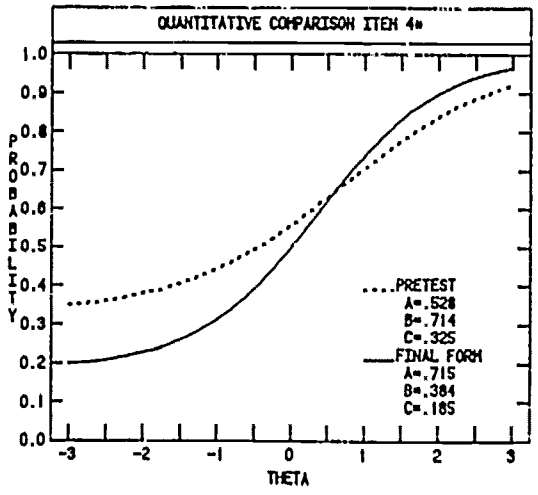
BEST COPY

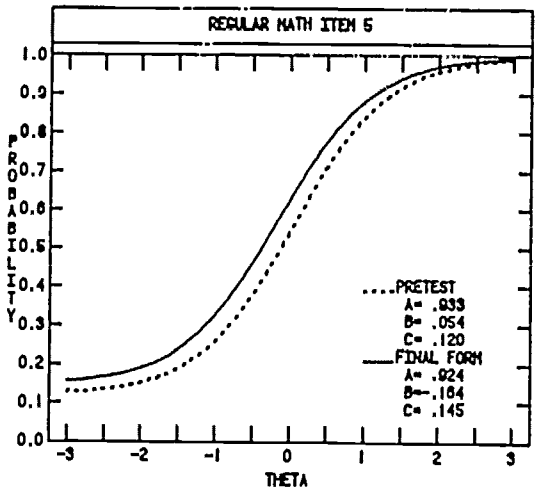
Figure 24: Plots of item response functions based on pretest and final form item parameter estimates for fourteen problematic Form 3BSA3 mathematical items.

*Item for which difference in pretest and final form item response functions can potentially be explained by item position effects.









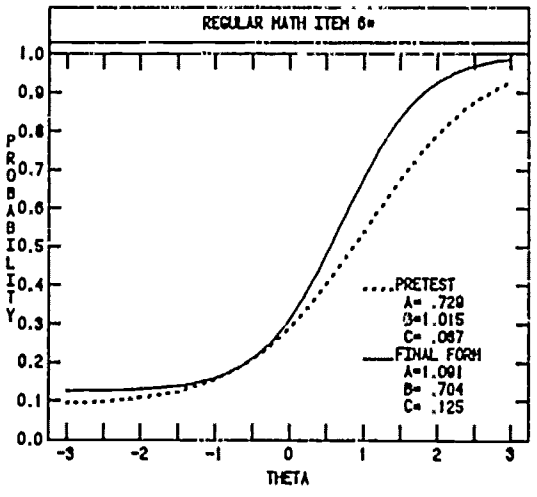


Figure 24: Plots of item response functions based on pretest and final form parameter estimates for fourteen problematic Form 3BSA3 mathematical items.

*Item for which difference in pretest and final form item response functions can potentially be explained by item position effects.

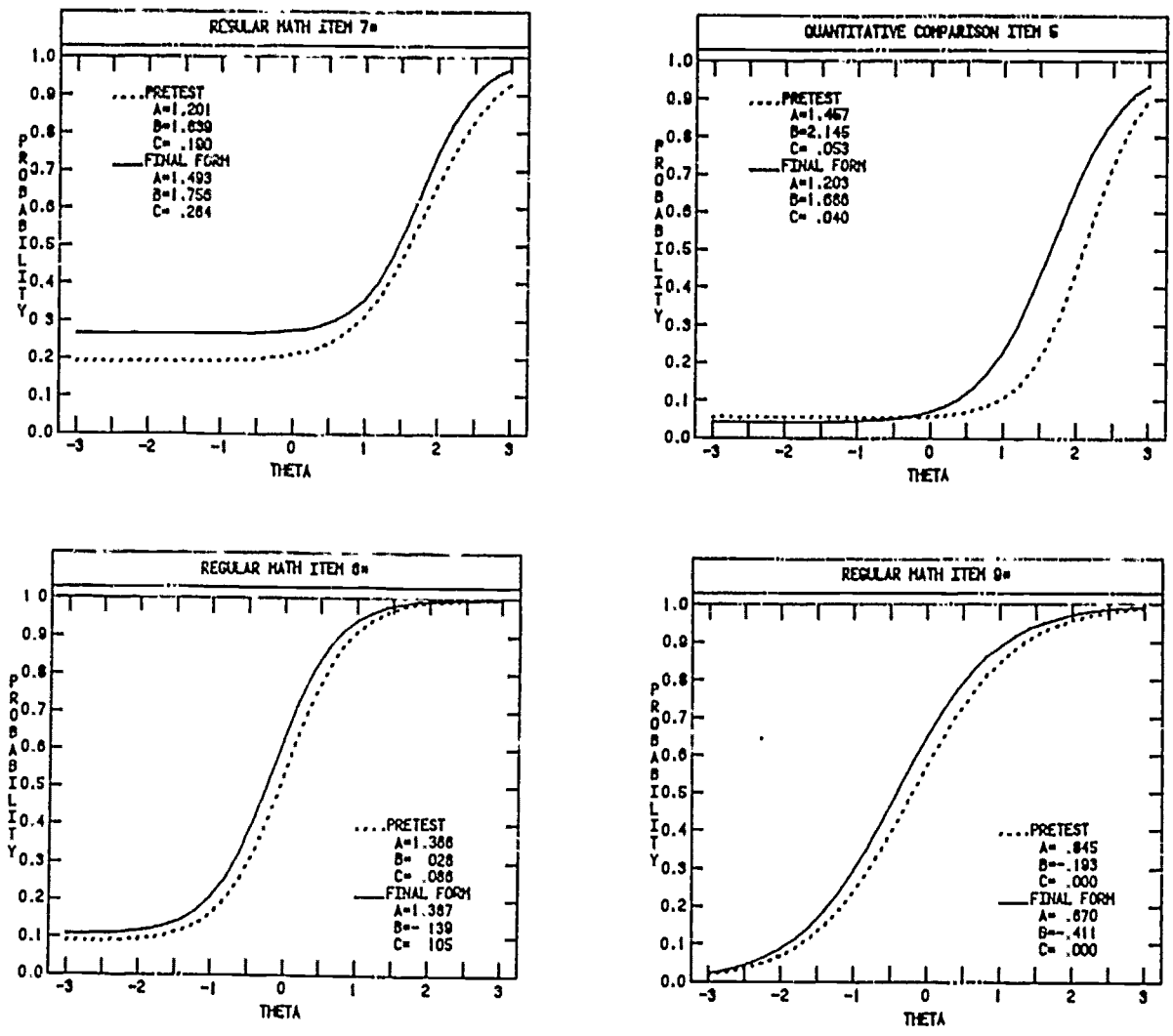


Figure 24: Plots of item response functions based on pretest and final form parameter estimates for fourteen problematic Form 3BSA3 mathematical items.

*Item for which difference in pretest and final form item response functions can potentially be explained by item position effects.

BEST COPY . . .

Table 9

Form 3ASA3 and 3BSA3 Items for Which Position Effects Can Be Used as an Explanation for Pretest-Final Form Differences in Item Response Functions and Item Difficulty Estimates

Form	Item Label in Figure 23 or 24	Position of Item in Pretest[1] and Final Form
3ASA3	Quantitative Comparison Item 2	At end of pretest section (item 33); in middle of final form section (item 22 of 35).
	Regular Math Item 3	First regular 5-choice item in cluster of five-choice items (item 28); in middle of final form section (item 9 of 25).
3BSA3	Regular Math Item 2	First regular 5-choice item in cluster of five-choice items (item 28); in middle of final form section (item 12 of 25).
	Quantitative Comparison Item 1	At end of pretest section (item 33); in middle of final form section (item 21 of 35).
	Quantitative Comparison Item 2	At end of pretest section (item 32); in middle of final form section (item 24 of 35).
	Quantitative Comparison Item 4	At beginning of quantitative comparison items in pretest (item 17 of quantitative cluster items 16-35); in middle of final form (item 15 of 35).
	Regular Math Item 6	At end of pretest section (item 35); near but not at the end of final form section (item 30 of 35).
	Regular Math Item 7	Second regular 5-choice item in cluster of five-choice items (item 29); in middle of final form section (item 11 of 25).
	Regular Math Item 8	At beginning of pretest section (item 3); in middle of final form (item 9 of 25).
	Regular Math Item 9	First item in pretest section (item 1); in middle of final form (item 11 of 25).

[1] In newer 35 item mathematical pretests, the five-choice regular mathematics items, of which there are 15 in total, are located in item positions 1-7 and 28-35, while the 20 four-choice quantitative comparison items are located in positions 8-27. For older mathematics pretests containing four-choice and five-choice items, positioning was not predetermined.

(45% of the total 13 regular math items flagged) and four were quantitative comparison items (36% of the total 11 quantitative comparison item flagged). It should be noted that of the total 120 items used in this study (40 quantitative comparison and 80 regular math items), a much larger proportion of the quantitative comparison items (11 of 40, or 27.5%) demonstrated large absolute differences in item response functions than did regular math items (13 of 80, or 16.3%). Kingston and Dorans (1982) obtained similar results for quantitative comparison items in their study. In sum, it would appear that an overriding explanation for the differences in the pre-equating and intact final form IRT equatings based on item position effects cannot be advanced for the mathematical portion of this study.

Additional Equating Results

In concluding the results section, one other outcome should be mentioned; this follows from a form by form comparison of the intact form calibration system IRT equating to the intact form direct link equating and, in the case of 3BSA3 verbal, the intact form concurrent IRT equating. It would appear, based upon the equatings done, that the equating is adequate when done through the indirect linking of the new and old forms used for equating via the overall calibration system. That is, even though in this situation the forms to be equated are linked, in some instances, indirectly through intervening LOGIST runs, and parameter estimates placed on a scale defined by the ability distribution of the sample taking a form not used in the equatings, the quality of the equatings are comparable to those resulting from either linking the new and old forms directly (direct link equating) or, in the case of 3BSA3 verbal, calibrating all data concurrently so that

new and old form parameter estimates are automatically on the same scale.

In a subsequent review of this study, it was pointed out that for only certain of the verbal forms to be equated was it the case that these forms were separated by intervening LOGIST runs, and in these cases, the maximum number of intervening runs was one. Hence, the present study design does not really simulate a model in which intact final form IRT equating might take place in the future. In a follow-up of this study (Eignor and Stocking, 1985), the intact form calibration system IRT equatings will be redone using parameter estimates for these forms that are separated by more than one intervening LOGIST run. For instance, for the intact form calibration system IRT equating of SAT-mathematical form 3ASA3 to XSA2, 3ASA3 item parameter estimates from LOGIST run 4 (see Figure 2, page 9) and XSA2 item parameter estimates from LOGIST run 7 will be used in the equating. In this case, the forms are separated by a number of intervening LOGIST runs, thereby simulating a more representative model of how intact form IRT equating of the SAT might take place.

Conclusions

The results of pre-equating the two forms of SAT-verbal reported on in this study, when compared to the intact final form IRT equatings, varied considerably, ranging from reasonably acceptable for Form 3ASA3 to only marginally acceptable or unacceptable for Form 3BSA3. Contributing reasons for the inferiority of the Form 3BSA3 pre-equating results, having to do with the location of reading passages and reading comprehension items at the

end of pretest sections, have been advanced and discussed. The verbal results reported here have clear implications for changes in test development practice, having to do with the positioning of pretest and final form reading comprehension items and the making of minor changes in the wording of items between pretest and final form, if pre-equating the SAT-verbal section is to be in any way a reality.

The results of pre-equating the two forms of SAT-mathematical, when compared to the relevant intact final form IRT equatings, were fairly similar and have to be considered only marginally acceptable, or perhaps, unacceptable. Unlike the pre-equating of Form 3BSA3 verbal, contributing reasons for the discrepant 3ASA3 and 3BSA3 mathematical pre-equatings could not be clearly advanced. For certain of the mathematical items demonstrating large differences in item response functions between pretest and final form, the positions of these items in the pretests could be offered as an explanation for the differences. For the other items demonstrating large differences, no explanation, other than there was a higher percentage of four-choice quantitative comparison items in this group, could be advanced.

For the three pre-equatings deemed marginally acceptable to unacceptable, perhaps of greater concern than the fact that a certain percentage of the items were estimated as clearly being more difficult in pretest than in final form (these were the items that were specifically discussed), is the fact that, for these three pre-equatings, an overwhelming percentage of the total number of items had higher difficulty estimates in pretest than in final form. For many of these items, the differences in the pretest - final form item difficulty estimates were slight, but considered

collectively, these difference are clearly a contributor to the poor pre-equating results. A clear explanation for this occurrence, given the design of this study, is difficult to advance. At least five potential explanations may be offered at this point, however. For certain of these potential explanations, data were available that allowed further investigation; the results are contained in this report. For other of the potential explanations, additional data analysis activities and subsequent investigation will be necessary. These potential explanations are as follows:

1. The design of this study is such that, for both SAT-verbal and SAT-mathematical, the first block of items, calibrated in LOGIST run 1 depicted in Figure 1 or 2, is connected to the last block of items by only a single chain of some 15 separate links. Each link involves LOGIST estimation and then a superimposed scaling or linking run. Any weakness in a particular link will be carried across all additional following links. A better design for this study would, perhaps, have been the placement of bridging cross-links that would have strengthened the overall linkages necessary in Figures 1 or 2. Financial considerations precluded the location and calibration of these cross-links in this study; they will be investigated, however, in a follow-up study (Eignor and Stocking, 1985).
2. The scaling procedure used in this study (Stocking and Lord, 1983) may not provide accurate linking. While the scaling procedure used has been well researched, it has not been applied consecutively as many times in past research studies to link separate LOGIST runs as it has been in this study. In the follow-up to this study (Eignor and Stocking, 1985), all the data for SAT-mathematical data depicted in Figure 2 of this report will be run in one large LOGIST run, thereby circumventing the need for the scaling runs, and the

pre-equating resulting from this calibration of the pretest items will then be compared to the criterion equating and the pre-equating results from this study. This should provide data upon which to evaluate whether the poor pre-equating results from the present study were the result of problems with the scaling procedure.

3. Certain mechanics of the item calibration process used in this study may have contributed to the discrepancies in the pretest and final form item difficulty parameter estimates and the resultant IRT equatings. For instance, would the difficulties of the pretest items have been different had the entire pretests been calibrated, and not just the specific pretest items needed for this study? Revisiting certain of the pretests, calibrating the entire pretest section, and then comparing the new parameter estimates for items of interest to the parameter estimates observed in this study, could provide an answer.

4. The discrepancies in the pretest and final form item difficulty estimates, and the resultant IRT equatings, may be due to context effects (i.e., the relationship between the item of interest and adjacent items), which because of the nature of the design of this study cannot readily be isolated. While it is reasonable to assume that the context in which an item occurs may affect the parameter estimates that result (see Yen, 1980), it is a bit more difficult to envision that these context effects would be predominately in the same direction, which would have to have been the case, at least in

terms of item difficulty parameter estimates, in this study. Also, a careful review by the author, both in pretest and final form, of all items identified as having widely discrepant item response functions in this study failed to locate any sort of readily apparent context effect.

5. The discrepancies in the pretest and final form item difficulty estimates, and the resultant IRT equatings, may be the result of differences in the ability levels of the groups used for calibration purposes. Theoretically, IRT item parameters are supposed to be independent of the ability level of the group used in the calibration process; in practice, this is not always the case, in particular for item difficulty estimates (Cook, Eignor, and Petersen, 1982). It can be hypothesized, if indeed the item difficulty parameter estimates are dependent on the ability level of the groups used in the calibration, that for three of the four forms under investigation, the ability levels of the groups taking the pretests should be consistently lower than the ability level of the group taking the intact final form. Tables 10 and 11 have been prepared to substantiate this hypothesis. Table 10 contains the total group scaled score means on each SAT-verbal final form that contained, as its variable section, a pretest which in turn contained items needed in this study. The data is presented separately for 3ASA3 verbal and 3BSA3 verbal; the scaled score means for these two final forms used for comparison purposes are also presented. The pretests in which the problematic items specifically discussed in this report are located are also identified. Table 11 contains comparable data for the two forms of SAT-mathematical. As can be seen from the summary indices at the bottom of these tables,

Table 10

Verbal Scaled Score Means for Pretest and Intact Final Form Total Groups from Which Samples Were Drawn for Pre-equating SAT-verbal Forms 3ASA3 and 3BSA3

Pretest Form	LOGIST[1] Run No.	3ASA3 Verbal Number of Problem Items	Scaled Score Mean
C167	1	3	398
C168	1		399
W4057	3		446
X2222	8		437
X2163	8		433
X2134	8	1	430
X2216	8		434
X2128	8	2	436
C237	10	4	419
C238	10	1	418
W5014	11		428
Z4125	12	1	438
Z4066	12		434
		12	

Intact Final Form	Scaled Score Mean
3ASA3	441[2]

Pretest Form	LOGIST[1] Run No.	3BSA3 Verbal Number of Problem Items	Scaled Score Mean
C167	1	1	398
C168	1		399
X4058	2		440
A1128	3		409
A2120	5	4	437
A2061	5	1	437
A5050	6		409
X5126	6		429
X5161	6		427
X5132	6	1	428
X5111	6	1	427
X2222	8		437
X2111	8		436
X2069	8		433
X2163	8		433
Z5069	10		410
C237	10	1	419
C238	10	5	418
		14	

Intact Final Form	Scaled Score Mean
3BSA3	431[2]

Number and percentage of times mean scaled score for pretest total group was less than mean scaled score for intact final form total group:

Form	Number	Percentage
3ASA3	12	92.3
3BSA3	11	61.1

[1]LOGIST run number refers to identification scheme in Figure 1.

[2]Scaled score mean resulting from application of linear parameters actually used to place form on scale operationally.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

Table 11

Mathematical Scaled Score Means for Pretest and Intact Final Form Total Groups from Which Samples Were Drawn for Pre-equating SAT-mathematical Forms 3ASA3 and 3BSA3

3ASA3 Mathematical

Pretest Form	LOGIST[1] Run No.	Number of Problem Items	Scaled Score Mean
C1613	3	1	441
C1614	3	2	442
X415	5		459
X316	6		496
X233	7		477
X241	7	2	477
X226	7		476
X232	7		479
X234	7		474
X243	7		476
Z515	10		448
Z512	10	1	444
C2314	14		452
C2318	14	3	455
		10	

Intact Final Form	Scaled Score Mean
3ASA3	485[2]

3BSA3 Mathematical

Pretest Form	LOGIST[1] Run No.	Number of Problem Items	Scaled Score Mean
W503	1		459
Z415	3	1	473
C1613	3	1	441
C1614	3	2	442
X413	5		462
X412	5	1	459
X313	6		492
X315	6	2	494
X233	7	1	478
X232	7		479
X235	7	1	477
X231	7		480
W305	9		493
Z515	10		448
Z512	10		444
X523	11	1	479
X521	11	1	404
X522	11		477
X525	11	1	478
Z203	13		462
C2314	14	2	452
C2318	14		455
		14	

Intact Final Form	Scaled Score Mean
3BSA3	478[2]

Number and percentage of times mean scaled score for pretest total group was less than mean scaled score for intact final form total group:

Form	Number	Percentage
3ASA3	13	92.9
3BSA3	13	59.1

[1] LOGIST run number refers to identification scheme in Figure 2.

[2] Scaled score mean resulting from application of linear parameters actually used to place form on scale operationally.

the hypothesis that the ability levels of the groups used in the calibration process may be influencing the resulting item difficulty estimates and equatings, while perhaps true for the problematic pre-equatings, is not borne out when one looks at 3ASA3 verbal, the only form for which the pre-equating was satisfactory. It may be the case, however, that an analysis at a more detailed level is necessary to address this issue. Unfortunately, the data presented in Tables 10 and 11 is the only data presently available to address this issue.

On a more general level, the results of this study also indicate that the IRT item parameter estimates generated for items given in pretest form do not remain invariant when given in intact final forms. This was true for the item discrimination and lower asymptote parameter estimates in this study although, unlike the item difficulty parameter estimates, this lack of invariance was not reflected in changes in parameter estimates in a specific direction. Based on the results of recent studies, particularly Cook, Eignor, and Petersen (1982), parameter invariance for all items in a test form would not be expected to be the case. Cook, et al, (1982) examined the stability over time of intact final form SAT-verbal and SAT-mathematical parameter estimates and the magnitudes of the discrepancies found in that study, based on the same intact final form given on two occasions, were of the magnitude of the discrepancies found for the item discrimination and lower asymptote parameter estimates in this study. The real issue is whether the lack of parameter invariance is serious enough to cause one to dismiss the use of item response theory for the particular application of concern. The application in this study is pre-equating, and the results of this study bring to serious question the feasibility of pre-equating the SAT. Further research, both in the direction of generating explanations for certain of the results of this study, and in the direction of replicating these results over other SAT forms, is clearly needed.

References

Angoff, W. H. Test reliability and effective test length. Psychometrika, 1953, 18, 1-14.

Angoff, W. H. Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), Educational Measurement (2nd ed.). Washington, DC: American Council on Education, 1971.

Bejar, I. I., and Wingersky, M. S. A study of pre-equating based on item response theory. Applied Psychological Measurement, 1982, 6, 309-325.

Cook, L. L., and Eignor, D. R. Practical considerations regarding the use of item response theory to equate tests. In R. K. Hambleton (Ed.), Applications of item response theory. Vancouver, B.C.: Educational Research Institute of British Columbia, 1983.

Cook, L. L., and Petersen, N. S. Item response theory equating for the SAT: New designs and directions. Proposal submitted to College Board--ETS Joint Staff Research and Development Committee, 1982.

Cook, L. L., Eignor, D. R., and Petersen, N. S. A study of the temporal stability of IRT item parameter estimates. A paper presented at the annual meeting of AERA, New York, 1982.

Eignor, D. R., and Stocking, M. L. IRT pre-equating: What went wrong? A paper to be presented at the annual meeting of AERA, Chicago, March 1985.

Kingston, N. M., and Dorans, N. J. The feasibility of using item response theory as a psychometric model for the GRE Aptitude Test. RR-82-12. Princeton, NJ: Educational Testing Service, 1982.

Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. Psychometrika, 1977, 14, 117-138.

Lord, F. M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum, 1980.

Petersen, N. S., Cook, L. L., and Stocking, M. L. IRT versus conventional equating methods: A comparative study of scale stability. Journal of Educational Statistics, 1983, 8, 137-156.

Stocking, M. L., and Lord, F. M. Developing a common metric in item response theory. Applied Psychological Measurement, 1983, 7, 201-210.

Wingersky, M. S. LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. K. Hambleton (Ed.), Applications of item response theory. Vancouver, B.C.: Educational Research Institute of British Columbia, 1983.

Wingersky, M. S., Barton, M. A., and Lord, F. M. LOGIST V user's guide. Princeton, NJ: Educational Testing Service, 1982.

Yen, W. M. The extent, causes, and importance of context effects on item parameters for two latent trait models. Journal of Educational Measurement, 1980, 17, 297-311.

Appendix

TABLE 1

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3ASA3 to Old Form XSA2

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
85	782.10	782.10	783.02	782.10
84	773.51	773.05	776.01	771.98
83	764.77	763.96	769.00	761.35
82	756.36	755.28	761.99	751.28
81	748.25	746.94	754.98	741.84
80	740.39	738.89	747.97	732.94
79	732.75	731.09	740.95	724.51
78	725.31	723.51	733.94	716.49
77	718.03	716.13	726.93	708.82
76	710.91	708.92	719.92	701.45
75	703.93	701.86	712.91	694.34
74	697.08	694.95	705.89	687.46
73	690.33	688.16	698.88	680.79
72	683.69	681.48	691.87	674.29
71	677.13	674.91	684.86	667.94
70	670.65	668.42	677.85	661.73
69	664.23	662.00	670.84	655.63
68	657.87	655.65	663.82	649.62
67	651.55	649.35	656.81	643.70
66	645.26	643.08	649.80	637.84
65	638.99	636.84	642.79	632.03
64	632.73	630.62	635.78	626.26
63	626.49	624.42	628.77	620.51
62	620.24	618.22	621.75	614.77
61	613.99	612.01	614.74	609.03
60	607.72	605.80	607.73	603.29
59	601.44	599.57	600.72	597.53
58	595.13	593.33	593.71	591.74
57	588.79	587.05	586.69	585.91
56	582.43	580.75	579.68	580.04
55	576.02	574.41	572.67	574.11
54	569.58	568.03	565.66	568.12
53	563.08	561.61	558.65	562.06
52	556.54	555.14	551.64	555.92
51	549.93	548.61	544.62	549.71
50	543.27	542.01	537.61	543.40
49	536.53	535.35	530.60	537.00
48	529.72	528.62	523.59	530.50
47	522.83	521.81	516.58	523.90
46	515.86	514.91	509.56	517.18
45	508.79	507.92	502.55	510.36
44	501.63	500.83	495.54	503.42
43	494.36	493.65	488.53	496.36
42	487.00	486.36	481.52	489.18
41	479.53	478.98	474.51	481.89
40	471.96	471.49	467.49	474.49
39	464.28	463.90	460.48	466.97
38	456.51	456.22	453.47	459.35
37	448.66	448.45	446.46	451.64
36	440.72	440.61	439.45	443.85
35	432.72	432.69	432.43	435.99
34	424.67	424.73	425.42	428.08
33	416.57	416.73	418.41	420.14
32	408.46	408.70	411.40	412.18

TABLE 1 (continued)

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3ASA3 to Old Form XSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
31	400.34	400.67	404.39	404.21
30	392.23	392.65	397.38	396.26
29	384.15	384.65	390.36	388.35
28	376.10	376.69	383.35	380.47
27	368.11	368.78	376.34	372.65
26	360.19	360.94	369.33	364.89
25	352.33	353.15	362.32	357.20
24	344.55	345.45	355.31	349.58
23	336.85	337.82	348.29	342.03
22	329.24	330.28	341.28	334.55
21	321.72	322.82	334.27	327.15
20	314.29	315.44	327.26	319.81
19	306.95	308.15	320.25	312.53
18	299.69	300.95	313.23	305.31
17	292.53	293.83	306.22	298.14
16	285.44	286.78	299.21	291.01
15	278.44	279.81	292.20	283.91
14	271.51	272.91	285.19	276.83
13	264.66	266.08	278.18	269.76
12	257.87	259.31	271.16	262.68
11	251.15	252.60	264.15	255.58
10	244.49	245.94	257.14	248.44
9	237.89	239.33	250.13	241.25
8	231.35	232.77	243.12	233.99
7	224.87	226.26	236.10	226.67
6	218.45	219.80	229.09	219.28
5	212.11	213.41	222.08	211.83
4	205.84	207.08	215.07	204.36
3	199.67	200.83	208.06	196.91
2	193.61	194.69	201.05	189.57
1	187.69	188.66	194.03	182.43
0	181.93	182.78	187.02	175.59
-1	176.34	177.07	180.01	169.12
-2	170.98	171.56	173.00	163.08
-3	165.86	166.28	165.99	157.43
-4	161.04	161.28	158.97	151.06
-5	156.71	156.74	151.96	144.29
-6	150.52	150.52	144.95	137.53
-7	143.39	143.39	137.94	130.76
-8	136.27	136.27	130.93	124.00
-9	129.14	129.14	123.92	117.23
-10	122.01	122.01	116.90	110.47
-11	114.88	114.88	109.89	103.70
-12	107.75	107.75	102.88	96.94
-13	100.63	100.63	95.87	90.17
-14	93.50	93.50	88.86	83.41
-15	86.37	86.37	81.85	76.64
-16	79.24	79.24	74.83	69.88
-17	72.11	72.11	67.82	63.12
-18	64.99	64.99	60.81	56.35
-19	57.86	57.86	53.80	49.59
-20	50.73	50.73	46.79	42.82
-21	43.60	43.60	39.77	36.06

TABLE 2

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3ASA3 to Old Form YSA3

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
85	774.04	774.04	798.78	774.04
84	768.88	768.83	791.50	768.14
83	762.71	762.60	784.23	760.67
82	756.14	755.99	776.96	752.78
81	749.38	749.19	769.68	744.87
80	742.55	742.32	762.41	737.10
79	735.71	735.46	755.14	729.55
78	728.92	728.65	747.86	722.23
77	722.19	721.90	740.59	715.15
76	715.55	715.24	733.32	708.27
75	708.98	708.66	726.05	701.58
74	702.48	702.16	718.77	695.05
73	696.05	695.72	711.50	688.66
72	689.67	689.33	704.23	682.38
71	683.33	682.99	696.95	676.21
70	677.02	676.68	689.68	670.11
69	670.73	670.38	682.41	664.07
68	664.43	664.09	675.13	658.08
67	658.14	657.80	667.86	652.12
66	651.83	651.49	660.59	646.18
65	645.49	645.16	653.32	640.24
64	639.13	638.80	646.04	634.30
63	632.72	632.40	638.77	628.35
62	626.27	625.96	631.50	622.37
61	619.78	619.48	624.22	616.35
60	613.22	612.93	616.95	610.28
59	606.61	606.33	609.68	604.16
58	599.94	599.67	602.40	597.98
57	593.20	592.95	595.13	591.72
56	586.40	586.16	587.86	585.39
55	579.52	579.29	580.59	578.96
54	572.57	572.36	573.31	572.45
53	565.55	565.35	566.04	565.84
52	558.45	558.27	558.77	559.13
51	551.28	551.11	551.49	552.32
50	544.03	543.88	544.22	545.43
49	536.71	536.57	536.95	538.38
48	529.32	529.20	529.67	531.26
47	521.86	521.76	522.40	524.04
46	514.34	514.25	515.13	516.73
45	506.77	506.70	507.86	509.33
44	499.14	499.09	500.58	501.85
43	491.47	491.44	493.31	494.31
42	483.77	483.75	486.04	486.71
41	476.04	476.03	478.76	479.06
40	468.28	468.30	471.49	471.38
39	460.52	460.55	464.22	463.67
38	452.75	452.80	456.94	455.94
37	444.98	445.04	449.67	448.22
36	437.22	437.30	442.40	440.49
35	429.47	429.57	435.13	432.78
34	421.73	421.85	427.85	425.09
33	414.01	414.14	420.58	417.42
32	406.31	406.45	413.31	409.78

TABLE 2 (continued)

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3ASA3 to Old Form YSA3

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
31	398.62	398.78	406.03	402.16
30	390.95	391.12	398.76	394.57
29	383.28	383.47	391.49	387.01
28	375.63	375.83	384.21	379.46
27	367.98	368.20	376.94	371.93
26	360.33	360.56	369.67	364.42
25	352.67	352.92	362.40	356.90
24	345.02	345.28	355.12	349.39
23	337.36	337.64	347.85	341.87
22	329.70	330.00	340.58	334.34
21	322.05	322.35	333.30	326.80
20	314.40	314.72	326.03	319.25
19	306.76	307.10	318.76	311.68
18	299.15	299.50	311.48	304.11
17	291.58	291.93	304.21	296.52
16	284.06	284.42	296.94	288.94
15	276.60	276.96	289.67	281.35
14	269.21	269.58	282.39	273.77
13	261.91	262.28	275.12	266.21
12	254.71	255.07	267.85	258.65
11	247.61	247.98	260.57	251.12
10	240.64	240.99	253.30	243.61
9	233.79	234.13	246.03	236.12
8	227.07	227.41	238.75	228.66
7	220.53	220.82	231.48	221.23
6	214.07	214.38	224.21	213.86
5	207.81	208.10	216.94	206.57
4	201.72	201.99	209.66	199.40
3	195.82	196.07	202.39	192.40
2	190.13	190.35	195.12	185.67
1	184.65	184.84	187.84	179.27
0	179.42	179.58	180.57	173.30
-1	174.45	174.58	173.30	167.83
-2	169.76	169.87	166.02	162.87
-3	165.40	165.47	158.75	158.45
-4	161.40	161.44	151.48	152.35
-5	157.94	157.95	144.21	145.50
-6	151.81	151.81	136.93	138.65
-7	144.59	144.59	129.66	131.79
-8	137.37	137.37	122.39	124.94
-9	130.15	130.15	115.11	118.09
-10	122.93	122.93	107.84	111.24
-11	115.71	115.71	100.57	104.38
-12	108.49	108.49	93.29	97.53
-13	101.27	101.27	86.02	90.68
-14	94.05	94.05	78.75	83.83
-15	86.83	86.83	71.48	76.97
-16	79.61	79.61	64.20	70.12
-17	72.39	72.39	56.93	63.27
-18	65.17	65.17	49.66	56.42
-19	57.94	57.94	42.38	49.56
-20	50.72	50.72	35.11	42.71
-21	43.50	43.50	27.84	35.86

TABLE 3

Raw Score to Scaled Score Transformations
SAT Verbal Form 3ASA3 to Old Forms XSA2 and YSA3

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
85	778.07	778.07	790.76	778.07
84	771.20	770.94	783.62	770.06
83	763.74	763.28	776.48	761.01
82	756.25	755.64	769.34	752.03
81	748.82	748.07	762.20	743.35
80	741.47	740.61	755. 6	735.02
79	734.23	733.28	747.92	727.03
78	727.11	726.08	740.78	719.36
77	720.11	719.01	733.64	711.98
76	713.23	712.08	726.50	704.86
75	706.46	705.26	719.36	697.96
74	699.76	698.55	712.22	691.26
73	693.19	691.94	705.08	684.72
72	686.68	685.41	697.94	678.34
71	680.23	678.95	690.80	672.07
70	673.84	672.55	683.66	665.92
69	667. 8	666.19	676.52	659.85
68	661..5	659.87	669.38	653.85
67	654.34	653.57	662.24	647.91
66	648.54	647.28	655.10	642.01
65	642.24	641.02	647.96	636.14
64	635.93	634.71	640..82	630.28
63	629.60	628.41	633.68	624.43
62	623.26	622.09	626.54	618.57
61	616.88	615.74	619.40	612.69
60	610.47	609.37	612.26	606.79
59	604.02	602.95	6 5.12	600.85
58	597.54	596.50	597.98	594.86
57	591.00	590.00	590.84	588.82
56	584.41	583.45	583.70	582.71
55	577.77	576.85	576.56	576.54
54	571.07	570.20	569.42	570.28
53	564.32	563.48	562.28	563.95
52	557.50	556.70	555.14	557.53
51	550.61	549.86	548.00	551.01
50	543.65	542.95	540.86	544.40
49	536.62	535.96	533.72	537.69
48	529.52	528.91	526.58	530.88
47	522.35	521.78	519.44	523.97
46	515.10	514.58	512.30	516.95
45	507.78	507.31	505.16	509.84
44	500.38	499.96	498.02	502.64
43	492.92	492.54	49..88	495.34
42	485.39	485.06	483.74	487.95
41	477.78	477.50	476.60	480.48
40	470.12	469.89	469.46	472.93
39	462.40	462.23	462.32	465.32
38	454.63	454.51	455.18	457.65
37	446.82	446.75	448.04	449.93
36	438.97	438.95	440.90	442.17
35	431.10	431.13	433.76	434.39
34	423.20	423.29	426.62	426.59
33	415 29	415.43	419.48	418.78

[1] nsformation used for operational score reporting purposes.

TABLE 3 (continued)

Raw Score to Scaled Score Transformations
SAT Verbal Form 3ASA3 to Old Forms XSA2 and YSA3

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
32	407.38	407.58	412.34	410.98
31	399.48	399.73	405.20	403.19
30	391.59	391.88	398.06	395.42
29	383.72	384.06	390.92	387.68
28	375.87	376.26	383.78	379.97
27	368.04	368.49	376.64	372.29
26	360.26	360.75	369.50	364.65
25	352.50	353.04	362.36	357.05
24	344.78	345.36	355.22	349.48
23	337.11	337.73	348.08	341.95
22	329.47	330.14	340.94	334.45
21	321.88	322.58	333.80	326.97
20	314.34	315.08	326.66	319.53
19	306.86	307.62	319.52	312.11
18	299.42	300.22	312.38	304.71
17	292.06	292.88	305.24	297.33
16	284.75	285.60	298.10	289.98
15	277.52	278.39	290.96	282.63
14	270.36	271.24	283.82	275.30
13	263.28	264.18	276.68	267.98
12	256.29	257.19	269.54	260.67
11	249.38	250.29	262.40	253.35
10	242.57	243.47	255.26	246.02
9	235.84	236.73	248.12	238.68
8	229.21	230.09	240.98	231.33
7	222.68	223.54	233.84	223.95
6	216.26	217.09	226.69	216.57
5	209.96	210.75	219.55	209.20
4	203.78	204.53	212.41	201.88
3	197.75	198.45	205.27	194.66
2	191.87	192.52	198.13	187.62
1	186.17	186.75	190.99	180.85
0	180.67	181.18	183.85	174.45
-1	175.39	175.82	176.71	168.47
-2	170.37	170.71	169.57	162.97
-3	165.63	165.88	162.43	157.94
-4	161.22	161.36	155.29	151.70
-5	157.32	157.34	148.15	144.90
-6	151.17	151.17	141.01	138.09
-7	143.99	143.99	133.87	131.28
-8	136.82	136.82	126.73	124.47
-9	129.64	129.64	119.59	117.66
-10	122.47	122.47	112.45	110.85
-11	115.30	115.30	105.31	104.04
-12	108.12	108.12	98.17	97.24
-13	100.95	100.95	91.03	90.43
-14	93.77	93.77	83.89	83.62
-15	86.60	86.60	76.75	76.81
-16	79.42	79.42	69.61	70.00
-17	72.25	72.25	62.47	63.19
-18	65.08	65.08	55.33	56.38
-19	57.90	57.90	48.19	49.58
-20	50.73	50.73	41.05	42.77
-21	43.55	43.55	33.91	35.96

[1]Transformation used for operational score reporting purposes.

TABLE 4

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Form YSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
85	801.48	801.51	768.54	801.51
84	791.12	791.32	761.68	793.20
83	782.48	783.21	754.82	787.37
82	773.75	775.16	747.96	781.48
81	765.08	767.17	741.11	775.42
80	756.51	759.22	734.25	769.16
79	748.08	751.30	727.39	762.70
78	739.77	743.40	720.53	756.01
77	731.57	735.50	713.67	749.14
76	723.46	727.59	706.81	742.08
75	715.42	719.65	699.95	734.84
74	707.42	711.68	693.09	727.44
73	699.45	703.68	686.24	719.89
72	691.50	695.65	679.38	712.21
71	683.56	687.57	672.52	704.41
70	675.61	679.46	665.66	696.50
69	667.65	671.32	658.80	688.50
68	659.69	663.15	651.94	680.41
67	651.73	654.97	645.08	672.26
66	643.76	646.77	638.23	664.07
65	635.80	638.58	631.37	655.84
64	627.85	630.40	624.51	647.59
63	619.93	622.24	617.65	639.34
62	612.04	614.12	610.79	631.10
61	604.19	606.03	603.93	622.89
60	596.40	598.00	597.07	614.72
59	588.66	590.03	590.21	606.61
58	580.99	582.14	583.36	598.56
57	573.40	574.31	576.50	590.58
56	565.88	566.58	569.64	582.68
55	558.45	558.93	562.78	574.86
54	551.10	551.37	555.92	567.12
53	543.84	543.91	549.06	559.47
52	536.66	536.55	542.20	551.91
51	529.56	529.27	535.34	544.43
50	522.53	522.09	528.49	537.04
49	515.59	515.00	521.63	529.73
48	508.70	507.99	514.77	522.50
47	501.88	501.06	507.91	515.35
46	495.11	494.21	501.05	508.26
45	488.40	487.42	494.19	501.24
44	481.72	480.69	487.33	494.27
43	475.07	474.01	480.48	487.36
42	468.46	467.37	473.62	480.48
41	461.86	460.77	466.76	473.64
40	455.27	454.20	459.90	466.82
39	448.70	447.65	453.04	460.02
38	442.12	441.10	446.18	453.23
37	435.54	434.57	439.32	446.44
36	428.95	428.03	432.46	439.64
35	422.35	421.48	425.61	432.82
34	415.73	414.93	418.75	425.97
33	409.09	408.35	411.89	419.10
32	402.43	401.75	405.03	412.18

TABLE 4 (continued)

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Form YSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
31	395.73	395.13	398.17	405.22
30	389.02	388.49	391.31	398.22
29	382.28	381.82	384.45	391.17
28	375.51	375.12	377.59	384.06
27	368.72	368.41	370.74	376.91
26	361.90	361.67	363.88	369.71
25	355.06	354.91	357.02	362.46
24	348.19	348.14	350.16	355.17
23	341.30	341.36	343.30	347.84
22	334.38	334.57	336.44	340.48
21	327.45	327.77	329.58	333.08
20	320.49	320.98	322.72	325.65
19	313.52	314.19	315.87	318.21
18	306.52	307.40	309.01	310.73
17	299.51	300.62	302.15	303.25
16	292.48	293.85	295.29	295.74
15	285.45	287.09	288.43	288.22
14	278.41	280.34	281.57	280.69
13	271.37	273.62	274.71	273.16
12	264.34	266.91	267.86	265.64
11	257.33	260.22	261.00	258.13
10	250.35	253.56	254.14	250.64
9	243.41	246.93	247.28	243.20
8	236.52	240.34	240.42	235.82
7	229.69	233.80	233.56	228.53
6	222.94	227.31	226.70	221.36
5	216.28	220.87	219.84	214.32
4	209.73	214.51	212.99	207.45
3	203.30	208.23	206.13	200.79
2	197.02	202.05	199.27	194.35
1	190.90	195.96	192.41	188.16
0	184.96	189.99	185.55	182.24
-1	179.22	184.14	178.69	176.61
-2	173.69	178.42	171.83	171.29
-3	168.38	172.84	164.98	166.28
-4	163.31	167.40	158.12	161.61
-5	158.47	162.09	151.26	157.31
-6	153.89	156.86	144.40	153.42
-7	149.37	151.56	137.54	150.11
-8	141.73	144.32	130.68	142.38
-9	134.10	136.62	123.82	134.66
-10	126.47	128.91	116.96	126.93
-11	118.83	121.20	110.11	119.20
-12	111.20	113.49	103.25	111.48
-13	103.57	105.78	96.39	103.75
-14	95.93	98.07	89.53	96.03
-15	88.30	90.36	82.67	88.30
-16	80.66	82.65	75.81	80.57
-17	73.03	74.94	68.95	72.85
-18	65.40	67.24	62.09	65.12
-19	57.76	59.53	55.24	57.39
-20	50.13	51.82	48.38	49.67
-21	42.50	44.11	41.52	41.94

TABLE 5

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Form 3ASA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
85	792.53	792.54	760.59	792.54
84	779.06	778.64	753.79	781.30
83	768.97	768.40	746.98	773.51
82	759.88	759.14	740.18	766.34
81	751.32	750.44	733.37	759.43
80	743.09	742.09	726.57	752.58
79	735.38	733.97	719.76	745.71
78	727.20	726.00	712.96	738.78
77	719.42	718.14	706.15	731.78
76	711.70	710.36	699.35	724.68
75	704.01	702.62	692.55	717.49
74	696.35	694.92	685.74	710.21
73	688.70	687.25	678.94	702.86
72	681.06	679.60	672.13	695.43
71	673.42	671.96	665.33	687.95
70	665.79	664.33	658.52	680.41
69	658.15	656.71	651.72	672.83
68	650.52	649.10	644.91	665.22
67	642.89	641.50	638.11	657.59
66	635.27	633.91	631.30	649.95
65	627.66	626.33	624.50	642.30
64	620.06	618.77	617.70	634.66
63	612.47	611.24	610.89	627.03
62	604.92	603.74	604.09	619.42
61	597.41	596.28	597.28	611.84
60	589.93	588.85	590.48	604.30
59	582.49	581.48	583.67	596.81
58	575.12	574.16	576.87	589.37
57	567.80	566.90	570.06	581.99
56	560.54	559.70	563.26	574.66
55	553.35	552.57	556.45	567.40
54	546.23	545.51	549.65	560.21
53	539.18	538.52	542.85	553.08
52	532.20	531.60	536.04	546.01
51	525.29	524.74	529.24	539.01
50	518.44	517.95	522.43	532.06
49	511.66	511.22	515.63	525.17
48	504.93	504.55	508.82	518.34
47	498.26	497.93	502.02	511.55
46	491.64	491.36	495.21	504.81
45	485.05	484.83	488.41	498.10
44	478.50	478.33	481.61	491.42
43	471.98	471.86	474.80	484.77
42	465.48	465.41	468.00	478.13
41	459.00	458.98	461.19	471.50
40	452.52	452.56	454.39	464.88
39	446.05	446.14	447.58	458.25
38	439.58	439.72	440.78	451.61
37	433.11	433.29	433.97	444.95
36	426.62	426.86	427.17	438.27
35	420.13	420.41	420.37	431.57
34	413.62	413.95	413.56	424.83
33	407.09	407.48	406.76	418.06
32	400.55	400.98	399.95	411.25

TABLE 5 (continued)

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Form 3ASA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
31	394.00	394.48	393.15	404.40
30	387.42	387.95	386.34	397.51
29	380.84	381.42	379.54	390.58
28	374.23	374.86	372.73	383.61
27	367.62	368.30	365.93	376.61
26	361.00	361.73	359.12	369.57
25	354.38	355.15	352.32	362.50
24	347.75	348.56	345.52	355.40
23	341.12	341.97	338.71	348.27
22	334.49	335.39	331.91	341.12
21	327.88	328.81	325.10	333.95
20	321.27	322.24	318.30	326.76
19	314.68	315.68	311.49	319.56
18	308.11	309.13	304.69	312.35
17	301.56	302.61	297.88	305.14
16	295.04	296.11	291.08	297.93
15	288.55	289.65	284.28	290.73
14	282.10	283.21	277.47	283.54
13	275.69	276.82	270.67	276.39
12	269.33	270.46	263.86	269.26
11	263.03	264.16	257.06	262.19
10	256.78	257.91	250.25	255.18
9	250.59	251.71	243.45	248.23
8	244.46	245.58	236.64	241.38
7	238.41	239.50	229.84	234.63
6	232.42	233.49	223.03	227.99
5	226.50	227.55	216.23	221.49
4	220.65	221.66	209.43	215.12
3	214.87	215.84	202.62	208.91
2	209.15	210.08	195.82	202.85
1	203.49	204.37	189.01	196.96
0	197.88	198.71	182.21	191.24
-1	192.32	193.08	175.40	185.67
-2	186.78	187.48	168.60	180.27
-3	181.25	181.86	161.79	175.01
-4	175.68	176.20	154.99	169.88
-5	169.99	170.42	148.19	164.84
-6	164.01	164.30	141.38	159.79
-7	156.85	156.96	134.58	153.60
-8	148.10	148.10	127.77	146.26
-9	140.78	140.78	120.97	138.92
-10	133.46	133.46	114.16	131.58
-11	126.14	126.14	107.36	124.24
-12	118.82	118.82	100.55	116.90
-13	111.49	111.49	93.75	109.57
-14	104.17	104.17	86.94	102.23
-15	96.85	96.85	80.14	94.89
-16	89.53	89.53	73.34	87.55
-17	82.20	82.20	66.53	80.21
-18	74.88	74.88	59.73	72.87
-19	67.56	67.56	52.92	65.53
-20	60.24	60.24	46.12	58.19
-21	52.91	52.91	39.31	50.86

TABLE 6

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Forms YSA2 and 3ASA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
85	797.01	797.03	764.55	797.03
84	785.09	785.01	757.72	787.25
83	775.72	775.81	750.89	780.44
82	766.81	767.15	744.06	773.91
81	758.20	758.81	737.22	767.43
80	749.80	750.66	730.39	760.87
79	741.58	742.64	723.56	754.20
78	733.49	734.70	716.73	747.40
77	725.49	726.82	709.90	740.46
76	717.58	718.97	703.07	733.38
75	709.71	711.14	696.24	726.16
74	701.88	703.30	689.40	718.83
73	694.08	695.47	682.57	711.38
72	686.28	687.62	675.74	703.82
71	678.49	679.76	668.91	696.18
70	670.70	671.89	662.08	688.45
69	662.90	664.01	655.25	680.66
68	655.11	656.12	648.41	672.82
67	647.31	648.23	641.58	664.93
66	639.51	640.34	634.75	657.01
65	631.73	632.46	627.92	649.07
64	623.96	624.58	621.09	641.12
63	616.20	616.74	614.26	633.18
62	608.48	608.93	607.43	625.26
61	600.80	601.15	600.59	617.37
60	593.16	593.43	593.76	609.51
59	585.58	585.75	586.93	601.71
58	578.05	578.15	580.11	593.97
57	570.60	570.61	573.27	586.28
56	563.21	563.14	566.44	578.67
55	555.90	555.75	559.61	571.13
54	548.67	548.44	552.77	563.66
53	541.51	541.21	545.94	556.27
52	534.43	534.07	539.11	548.96
51	527.42	527.01	532.28	541.72
50	520.49	520.02	525.45	534.55
49	513.62	513.11	518.62	527.45
48	506.82	506.27	511.78	520.42
47	500.07	499.50	504.95	513.45
46	493.37	492.79	498.12	506.53
45	486.72	486.12	491.29	499.67
44	480.11	479.51	484.46	492.85
43	473.53	472.94	477.63	486.06
42	466.97	466.39	470.80	479.30
41	460.43	459.88	463.96	472.57
40	453.90	453.38	457.13	465.85
39	447.37	446.89	450.30	459.13
38	440.85	440.41	443.47	452.42
37	434.32	433.93	436.64	445.70
36	427.79	427.44	429.81	438.96
35	421.24	420.95	422.98	432.19
34	414.67	414.44	416.14	425.40
33	408.09	407.91	409.31	418.58
32	401.49	401.37	402.48	411.72

[1] Transformation used for operational score reporting purposes.

TABLE 6 (continued)

Raw Score to Scaled Score Transformations
SAT-Verbal Form 3BSA3 to Old Forms YSA2 and 3ASA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
31	394.87	394.81	395.65	404.81
30	388.22	388.22	388.82	397.87
29	381.56	381.62	381.99	390.87
28	374.87	374.99	375.15	383.84
27	368.17	368.35	368.32	376.76
26	361.45	361.70	361.49	369.64
25	354.72	355.03	354.66	362.48
24	347.97	348.35	347.83	355.28
23	341.21	341.66	341.00	348.06
22	334.44	334.98	334.17	340.80
21	327.66	328.29	327.33	333.51
20	320.88	321.61	320.50	326.21
19	314.10	314.93	313.67	318.89
18	307.31	308.27	306.84	311.54
17	300.53	301.62	300.01	304.19
16	293.76	294.98	293.18	296.83
15	287.00	288.37	286.35	289.47
14	280.25	281.78	279.51	282.12
13	273.53	275.22	272.68	274.77
12	266.84	268.69	265.85	267.45
11	260.18	262.19	259.02	260.16
10	253.56	255.73	252.19	252.91
9	247.00	249.32	245.36	245.72
8	240.49	242.96	238.52	238.60
7	234.05	236.65	231.69	231.58
6	227.68	230.40	224.86	224.67
5	221.39	224.21	218.03	217.90
4	215.19	218.09	211.20	211.29
3	209.08	212.04	204.37	204.85
2	203.08	206.06	197.54	198.60
1	197.20	200.17	190.70	192.56
0	191.42	194.35	183.87	186.74
-1	185.77	188.61	177.04	181.14
-2	180.23	182.95	170.21	175.78
-3	174.81	177.35	163.38	170.64
-4	169.49	171.80	156.55	165.74
-5	164.23	166.25	149.72	161.07
-6	158.95	160.58	142.88	156.60
-7	153.11	154.26	136.05	151.85
-8	144.92	146.21	129.22	144.32
-9	137.44	138.70	122.39	136.79
-10	129.96	131.18	115.56	129.26
-11	122.49	123.67	108.73	121.72
-12	115.01	116.15	101.89	114.19
-13	107.53	108.64	95.06	106.66
-14	100.05	101.12	88.23	99.13
-15	92.57	93.61	81.40	91.59
-16	85.10	86.09	74.57	84.06
-17	77.62	78.57	67.74	76.53
-18	70.14	71.06	60.91	69.00
-19	62.66	63.54	54.07	61.46
-20	55.18	56.03	47.24	53.93
-21	47.71	48.51	40.41	46.40

[1] Transformation used for operational score reporting purposes.

TABLE 7

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Form XSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
60	783.77	783.77	762.57	783.77
59	773.90	774.11	754.00	779.20
58	762.69	763.05	745.43	772.09
57	751.21	751.70	736.86	763.63
56	739.84	740.41	728.30	754.48
55	728.75	729.38	719.73	745.01
54	718.04	718.69	711.16	735.41
53	707.70	708.37	702.59	725.80
52	697.71	698.39	694.02	716.24
51	688.05	688.72	685.45	706.77
50	678.68	679.33	676.88	697.40
49	669.54	670.17	668.32	688.17
48	660.61	661.22	659.75	679.06
47	651.84	652.43	651.18	670.07
46	643.20	643.76	642.61	661.19
45	634.66	635.18	634.04	652.40
44	626.18	626.67	625.47	643.68
43	617.74	618.20	616.90	635.01
42	609.31	609.74	608.33	626.36
41	600.88	601.28	599.77	617.71
40	592.43	592.79	591.20	609.04
39	583.94	584.27	582.63	600.32
38	575.41	575.71	574.06	591.54
37	566.83	567.09	565.49	582.70
36	558.18	558.41	556.92	573.78
35	549.47	549.66	548.35	564.79
34	540.70	540.85	539.78	555.73
33	531.85	531.96	531.22	546.61
32	522.94	523.01	522.65	537.44
31	513.97	514.00	514.08	528.24
30	504.95	504.94	505.51	519.01
29	495.88	495.83	496.94	509.77
28	486.79	486.70	488.37	500.54
27	477.68	477.55	479.80	491.34
26	468.58	468.41	471.24	482.16
25	459.50	459.29	462.67	473.02
24	450.46	450.21	454.11	463.94
23	441.48	441.19	445.53	454.90
22	432.57	432.25	436.96	445.93
21	423.75	423.39	428.39	437.02
20	415.02	414.64	419.82	428.17
19	406.41	405.99	411.25	419.39
18	397.90	397.46	402.69	410.68
17	389.51	389.05	394.12	402.04
16	381.24	380.75	385.55	393.46
15	373.07	372.56	376.98	384.94
14	365.01	364.49	368.41	376.49
13	357.05	356.51	359.84	368.11
12	349.18	348.62	351.27	359.79
11	341.38	340.82	342.70	351.54
10	333.65	333.08	334.14	343.34
9	325.97	325.40	325.57	335.20
8	318.33	317.76	317.00	327.10
7	310.72	310.14	308.43	319.03

TABLE 7 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Form XSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
6	303.11	302.54	299.86	310.96
5	295.48	294.91	291.29	302.88
4	287.81	287.25	282.72	294.73
3	280.06	279.51	274.16	286.48
2	272.20	271.67	265.59	278.06
1	264.19	263.67	257.02	269.41
0	255.99	255.50	248.45	260.47
-1	247.57	247.11	239.88	251.19
-2	238.92	238.51	231.31	241.58
-3	230.10	229.75	222.74	231.80
-4	221.23	220.97	214.17	222.20
-5	212.62	212.46	205.61	213.35
-6	205.01	205.00	197.04	206.17
-7	196.93	196.93	188.47	199.17
-8	188.59	188.59	179.90	190.63
-9	180.24	180.24	171.33	182.08
-10	171.89	171.89	162.76	173.54
-11	163.55	163.55	154.19	164.99
-12	155.20	155.20	145.62	156.45
-13	146.86	146.86	137.06	147.91
-14	138.51	138.51	128.49	139.36
-15	130.17	130.17	119.92	130.82
-16	121.82	121.82	111.35	122.27
-17	113.48	113.48	102.78	113.73

TABLE 8

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Form ZSA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
60	785.82	785.82	782.14	785.82
59	776.57	776.92	773.46	779.83
58	767.38	767.97	764.78	772.42
57	758.09	758.89	756.10	764.39
56	748.68	749.63	747.42	756.08
55	739.19	740.25	738.75	747.62
54	729.67	730.83	730.07	739.05
53	720.19	721.41	721.39	730.39
52	710.79	712.04	712.71	721.68
51	701.51	702.77	704.03	712.94
50	692.36	693.61	695.35	704.20
49	683.34	684.56	686.68	695.48
48	674.45	675.63	678.00	686.81
47	665.67	666.80	669.32	678.18
46	656.99	658.07	660.64	669.61
45	648.38	649.41	651.96	661.10
44	639.84	640.80	643.28	652.63
43	631.34	632.23	634.61	644.19
42	622.86	623.69	625.93	635.78
41	614.40	615.16	617.25	627.37
40	605.94	606.63	608.57	618.95
39	597.47	598.09	599.89	610.50
38	588.99	589.53	591.21	602.00
37	580.49	580.96	582.54	593.46
36	571.97	572.36	573.86	584.87
35	563.44	563.75	565.18	576.21
34	554.88	555.12	556.50	567.50
33	546.31	546.47	547.82	558.74
32	537.73	537.81	539.14	549.93
31	529.14	529.14	530.47	541.09
30	520.53	520.45	521.79	532.23
29	511.93	511.77	513.11	523.35
28	503.32	503.08	504.43	514.48
27	494.72	494.40	495.75	505.61
26	486.13	485.73	487.07	496.76
25	477.55	477.07	478.40	487.93
24	468.98	468.43	469.72	479.13
23	460.44	459.81	461.04	470.36
22	451.91	451.21	452.36	461.63
21	443.41	442.64	443.68	452.94
20	434.92	434.09	435.00	444.27
19	426.46	425.56	426.33	435.64
18	418.01	417.05	417.65	427.04
17	409.58	408.56	408.97	418.45
16	401.16	400.08	400.29	409.87
15	392.74	391.61	391.61	401.30
14	384.32	383.15	382.94	392.74
13	375.91	374.70	374.26	384.17
12	367.51	366.25	365.58	375.60
11	359.10	357.82	356.90	367.03
10	350.71	349.41	348.22	358.46
9	342.33	341.01	339.54	349.89
8	333.96	332.64	330.87	341.33
7	325.61	324.29	322.19	332.76

TABLE 8 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Form ZSA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
6	317.28	315.96	313.51	324.20
5	308.96	307.66	304.83	315.62
4	300.64	299.37	296.15	307.00
3	292.31	291.08	287.47	298.33
2	283.95	282.77	278.80	289.57
1	275.54	274.41	270.12	280.65
0	267.05	265.99	261.44	271.54
-1	258.46	257.49	252.76	262.19
-2	249.80	248.93	244.08	252.59
-3	241.12	240.38	235.40	242.81
-4	232.55	231.97	226.73	233.10
-5	224.30	223.91	218.05	223.79
-6	216.63	216.47	209.37	215.32
-7	209.23	209.23	200.69	207.46
-8	200.77	200.77	192.01	198.95
-9	192.30	192.30	183.33	190.45
-10	183.84	183.84	174.66	181.94
-11	175.38	175.38	165.98	173.43
-12	166.92	166.92	157.30	164.92
-13	158.45	158.45	148.62	156.41
-14	149.99	149.99	139.94	147.90
-15	141.53	141.53	131.26	139.39
-16	133.07	133.07	122.59	130.89
-17	124.60	124.60	113.91	122.38

TABLE 9

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Forms XSA2 and ZSA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
60	784.79	784.79	772.30	784.79
59	775.24	775.51	763.67	779.51
58	765.03	765.51	755.05	772.26
57	754.65	755.29	746.42	764.01
56	744.26	745.02	737.80	755.28
55	733.97	734.82	729.18	746.32
54	723.85	724.76	720.56	737.23
53	713.94	714.89	711.93	728.10
52	704.25	705.21	703.31	718.96
51	694.78	695.74	694.69	709.85
50	685.52	686.47	686.06	700.80
49	676.44	677.37	677.44	691.83
48	667.53	668.42	668.82	682.93
47	658.75	659.61	660.19	674.13
46	650.09	650.91	651.57	665.40
45	641.52	642.29	642.95	656.75
44	633.01	633.73	634.32	648.16
43	624.54	625.21	625.70	639.60
42	616.09	616.72	617.08	631.07
41	607.64	608.22	608.45	622.54
40	599.18	599.71	599.83	613.99
39	590.71	591.18	591.21	605.41
38	582.20	582.62	582.58	596.77
37	573.66	574.02	573.96	588.08
36	565.08	565.39	565.34	579.32
35	556.45	556.71	556.72	570.50
34	547.79	547.98	548.09	561.62
33	539.08	539.22	539.47	552.68
32	530.33	530.41	530.85	543.69
31	521.55	521.57	522.22	534.66
30	512.74	512.70	513.60	525.62
29	503.91	503.80	504.98	516.56
28	495.06	494.89	496.35	507.51
27	486.20	485.98	487.73	498.47
26	477.36	477.07	479.11	489.46
25	468.52	468.18	470.48	480.47
24	459.72	459.32	461.86	471.53
23	450.96	450.50	453.24	462.63
22	442.24	441.73	444.61	453.78
21	433.58	433.02	435.99	444.98
20	424.97	424.36	427.37	436.22
19	416.43	415.78	418.74	427.52
18	407.96	407.26	410.12	418.86
17	399.55	398.80	401.50	410.24
16	391.20	390.41	392.87	401.67
15	382.91	382.09	384.25	393.12
14	374.67	373.82	375.63	384.62
13	366.48	365.60	367.00	376.14
12	358.34	357.44	358.38	367.70
11	350.24	349.32	349.76	359.29
10	342.18	341.24	341.14	350.90
9	334.15	333.20	332.51	342.55
8	326.15	325.20	323.89	334.21
7	318.16	317.21	315.27	325.90

[1]Transformation used for operational score reporting purposes.

TABLE 9 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3ASA3 to Old Forms XSA2 and ZSA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
6	310.19	309.25	306.64	317.58
5	302.22	301.29	298.02	309.25
4	294.23	293.31	289.40	300.87
3	286.19	285.30	280.77	292.41
2	278.08	277.22	272.15	283.81
1	269.87	269.04	263.53	275.03
0	261.52	260.74	254.90	266.01
-1	253.01	252.30	246.28	256.69
-2	244.36	243.72	237.66	247.08
-3	235.61	235.06	229.03	237.31
-4	226.89	226.47	220.41	227.65
-5	218.46	218.19	211.79	218.57
-6	210.82	210.73	203.16	210.74
-7	203.08	203.08	194.54	203.32
-8	194.68	194.68	185.92	194.79
-9	186.27	186.27	177.29	186.26
-10	177.87	177.87	168.67	177.74
-11	169.46	169.46	160.05	169.21
-12	161.06	161.06	151.42	160.69
-13	152.66	152.66	142.80	152.16
-14	144.25	144.25	134.18	143.63
-15	135.85	135.85	125.56	135.11
-16	127.44	127.44	116.93	126.58
-17	119.04	119.04	108.31	118.05

[1] Transformation used for operational score reporting purposes.

TABLE 10

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Form YSA2

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
60	782.98	782.98	780.27	782.98
59	776.90	776.91	771.73	779.28
58	769.90	769.93	763.19	776.24
57	761.93	761.97	754.65	771.49
56	753.35	753.40	746.11	765.48
55	744.35	744.42	737.57	758.87
54	735.11	735.18	729.03	751.87
53	725.76	725.84	720.48	744.56
52	716.40	716.48	711.94	737.01
51	707.12	707.20	703.40	729.26
50	697.95	698.03	694.86	721.36
49	688.92	689.00	686.32	713.33
48	680.02	680.10	677.78	705.19
47	671.25	671.32	669.24	696.97
46	662.59	662.66	660.70	688.66
45	654.03	654.10	652.16	680.25
44	645.54	645.61	643.62	671.76
43	637.12	637.19	635.08	663.16
42	628.75	628.80	626.54	654.47
41	620.39	620.44	618.00	645.67
40	612.04	612.09	609.46	636.78
39	603.69	603.73	600.92	627.80
38	595.30	595.34	592.37	618.74
37	586.88	586.92	583.83	609.59
36	578.42	578.45	575.29	600.37
35	569.90	569.93	566.75	591.09
34	561.33	561.35	558.21	581.75
33	552.71	552.72	549.67	572.37
32	544.03	544.04	541.13	562.96
31	535.31	535.31	532.59	553.52
30	526.54	526.54	524.05	544.08
29	517.74	517.73	515.51	534.64
28	508.90	508.89	506.97	525.20
27	500.04	500.02	498.43	515.76
26	491.16	491.13	489.89	506.31
25	482.26	482.23	481.35	496.85
24	473.35	473.31	472.80	487.38
23	464.43	464.38	464.26	477.87
22	455.50	455.45	455.72	468.33
21	446.58	446.53	447.18	458.74
20	437.66	437.60	438.64	449.11
19	428.76	428.69	430.10	439.43
18	419.87	419.80	421.56	429.70
17	411.01	410.93	413.02	419.92
16	402.16	402.09	404.48	410.11
15	393.35	393.27	395.94	400.26
14	384.57	384.48	387.40	390.40
13	375.82	375.73	378.86	380.52
12	367.10	367.01	370.32	370.65
11	358.42	358.33	361.78	360.79
10	349.78	349.68	353.23	350.97
9	341.18	341.08	344.69	341.19
8	332.62	332.52	336.15	331.47
7	324.10	324.00	327.61	321.81

TABLE 10 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Form YSA2

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
6	315.62	315.53	319.07	312.25
5	307.19	307.10	310.53	302.77
4	298.81	298.72	301.99	293.41
3	290.48	290.40	293.45	284.17
2	282.21	282.12	284.91	275.10
1	273.98	273.90	276.37	266.23
	265.81	265.74	267.83	257.63
-1	257.70	257.64	259.29	249.40
-2	249.66	249.60	250.75	241.62
-3	241.68	241.63	242.21	234.39
-4	233.76	233.72	233.66	227.74
-5	225.84	225.82	225.12	221.66
-6	217.86	217.85	216.58	216.12
-7	209.20	209.20	208.04	210.98
-8	200.67	200.67	199.50	202.08
-9	192.15	192.15	190.96	193.33
-10	183.62	183.62	182.42	184.59
-11	175.10	175.10	173.88	175.84
-12	166.57	166.57	165.34	167.09
-13	158.04	158.04	156.80	158.34
-14	149.52	149.52	148.26	149.60
-15	140.99	140.99	139.72	140.85
-16	132.47	132.47	131.18	132.10
-17	123.94	123.94	122.64	123.35

TABLE 11

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Form 3ASA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
60	809.37	809.37	783.39	809.37
59	797.87	797.97	774.79	803.29
58	787.77	787.93	766.18	798.30
57	777.71	777.92	757.57	791.91
56	767.64	767.89	748.97	782.37
55	757.56	757.86	740.36	773.43
54	747.50	747.82	731.75	764.39
53	737.46	737.81	723.15	755.28
52	727.46	727.83	714.54	746.12
51	717.53	717.91	705.94	736.88
50	707.67	708.06	697.33	727.57
49	697.91	698.29	688.72	718.18
48	688.25	688.62	680.12	708.72
47	678.67	679.04	671.51	699.20
46	669.19	669.54	662.90	689.59
45	659.80	660.14	654.30	679.91
44	650.48	650.80	645.69	670.15
43	641.25	641.55	637.08	660.31
42	632.08	632.36	628.48	650.41
41	622.98	623.24	619.87	640.45
40	613.94	614.17	611.26	630.45
39	604.95	605.17	602.66	620.42
38	596.04	596.23	594.05	610.39
37	587.18	587.35	585.45	600.38
36	578.38	578.52	576.84	590.41
35	569.66	569.78	568.23	580.50
34	561.00	561.10	559.63	570.68
33	552.41	552.49	551.02	560.96
32	543.89	543.95	542.41	551.36
31	535.44	535.48	533.81	541.90
30	527.07	527.09	525.20	532.57
29	518.76	518.76	516.59	523.38
28	510.52	510.50	507.99	514.32
27	502.34	502.30	499.38	505.39
26	494.22	494.16	490.78	496.58
25	486.14	486.06	482.17	487.85
24	478.11	478.01	473.56	479.21
23	470.10	469.99	464.96	470.62
22	462.13	461.99	456.35	462.08
21	454.16	454.01	447.74	453.56
20	446.20	446.03	439.14	445.05
19	438.25	438.05	430.53	436.54
18	430.28	430.07	421.92	428.00
17	422.30	422.07	413.32	419.44
16	414.29	414.05	404.71	410.84
15	406.25	405.99	396.10	402.18
14	398.17	397.90	387.50	393.47
13	390.05	389.76	378.89	384.69
12	381.87	381.57	370.29	375.85
11	373.63	373.32	361.68	366.93
10	365.32	365.00	353.07	357.92
9	356.94	356.60	344.47	348.84
8	348.47	348.12	335.86	339.68
7	339.91	339.56	327.25	330.44

TABLE 11 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Form 3ASA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear	IRT Pre-equating
6	331.26	330.90	318.65	321.11
5	322.51	322.16	310.04	311.71
4	313.67	313.32	301.43	302.24
3	304.73	304.38	292.83	292.70
2	295.67	295.33	284.22	283.12
1	286.51	286.18	275.62	273.56
	277.24	276.93	267.01	264.12
-1	267.87	267.59	258.40	254.99
-2	258.45	258.20	249.80	246.38
-3	249.05	248.83	241.19	238.52
-4	239.81	239.64	232.58	231.61
-5	230.97	230.86	223.98	225.73
-6	222.89	222.83	215.37	220.97
-7	215.31	215.31	206.76	217.73
-8	205.48	205.48	198.16	207.89
-9	195.65	195.65	189.55	197.76
-10	185.82	185.82	180.94	187.63
-11	175.99	175.99	172.34	177.50
-12	166.16	166.16	163.73	167.36
-13	156.34	156.34	155.13	157.23
-14	146.51	146.51	146.52	147.10
-15	136.68	136.68	137.91	136.97
-16	126.85	126.85	129.31	126.84
-17	117.02	117.02	120.70	116.71

TABLE 12

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Forms YSA2 and 3ASA1

Raw Score	Estimated Scaled Score: IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
60	796.17	796.17	781.82	796.17
59	787.39	787.44	773.25	791.28
58	778.84	778.93	764.68	787.27
57	769.82	769.95	756.10	781.25
56	760.49	760.65	747.53	773.92
55	750.96	751.14	738.96	766.15
54	741.30	741.50	730.38	758.13
53	731.61	731.82	721.81	749.92
52	721.93	722.16	713.24	741.57
51	712.32	712.55	704.66	733.07
50	702.81	703.04	696.09	724.46
49	693.41	693.65	687.52	715.75
48	684.13	684.36	678.94	706.96
47	674.96	675.18	670.37	698.08
46	665.89	666.10	661.80	689.12
45	656.91	657.12	653.22	680.08
44	648.01	648.21	644.65	670.95
43	639.19	639.37	636.08	661.74
42	630.41	630.58	627.50	652.44
41	621.68	621.84	618.93	643.06
40	612.99	613.13	610.36	633.62
39	604.32	604.45	601.78	624.11
38	595.67	595.79	593.21	614.56
37	587.03	587.13	584.64	604.98
36	578.40	578.49	576.06	595.39
35	569.78	569.85	567.49	585.79
34	561.16	561.23	558.92	576.21
33	552.56	552.60	550.34	566.66
32	543.96	543.99	541.77	557.16
31	535.37	535.39	533.20	547.71
30	526.80	526.81	524.62	538.33
29	518.25	518.24	516.05	529.01
28	509.71	509.69	507.48	519.76
27	501.19	501.16	498.90	510.57
26	492.69	492.65	490.33	501.44
25	484.20	484.15	481.75	492.35
24	475.73	475.66	473.18	483.29
23	467.27	467.19	464.61	474.25
22	458.81	458.72	456.03	465.20
21	450.37	450.27	447.46	456.15
20	441.93	441.82	438.89	447.08
19	433.50	433.37	430.31	437.98
18	425.08	424.94	421.74	428.85
17	416.65	416.50	413.17	419.68
16	408.23	408.07	404.59	410.47
15	399.80	399.63	396.02	401.22
14	391.37	391.19	387.45	391.93
13	382.93	382.74	378.87	382.61
12	374.49	374.29	370.30	373.25
11	366.03	365.82	361.73	363.86
10	357.55	357.34	353.15	354.45
9	349.06	348.84	344.58	345.02
8	340.54	340.32	336.01	335.57
7	332.00	331.78	327.42	326.13

[1] Transformation used for operational score reporting purposes.

TABLE 12 (continued)

Raw Score to Scaled Score Transformations
SAT-Math Form 3BSA3 to Old Forms YSA2 and 3ASA1

Raw Score	Estimated Scaled Score			
	IRT Intact Form (Direct Link)	IRT Intact Form (Calibration System)	Intact Form Linear[1]	IRT Pre-equating
6	323.44	323.22	318.86	316.68
5	314.85	314.63	310.29	307.24
4	306.24	306.02	301.71	297.82
3	297.60	297.39	293.14	288.44
2	288.94	288.73	284.57	279.11
1	280.25	280.04	275.99	269.89
	271.53	271.34	267.42	260.88
-1	262.79	262.61	258.85	252.19
-2	254.06	253.90	250.27	244.00
-3	245.37	245.23	241.70	236.46
-4	236.78	236.68	233.13	229.67
-5	228.41	228.34	224.55	223.69
-6	220.37	220.34	215.98	218.55
-7	212.25	212.25	207.41	214.35
-8	203.08	203.08	198.83	204.99
-9	193.90	193.90	190.26	195.55
-10	184.72	184.72	181.69	186.11
-11	175.54	175.54	173.11	176.67
-12	166.37	166.37	164.54	167.23
-13	157.19	157.19	155.97	157.79
-14	148.01	148.01	147.39	148.35
-15	138.84	138.84	138.82	138.91
-16	129.66	129.66	130.25	129.47
-17	120.48	120.48	121.67	120.03

[1]Transformation used for operational score reporting purposes.